THOMSON

Probabilidade e Estatística para Engenharia e Ciências

Jay L. Devore

1

Visão Geral e Estatística Descritiva

Introdução

Os conceitos e métodos estatísticos não são apenas úteis, como também indispensáveis na compreensão do mundo ao nosso redor. Eles fornecem meios de obtenção de novas percepções no que diz respeito ao comportamento de diversos fenômenos que você encontrará em seu campo de especialização em engenharia ou ciência.

A disciplina estatística nos ensina a fazer julgamentos inteligentes e a tomar decisões na presença de incertezas e variações. Sem incertezas ou variações, haveria pouca necessidade de estatísticos ou métodos estatísticos. Se cada componente de um determinado tipo tivesse exatamente o mesmo tempo de vida, se todos os resistores produzidos por um determinado fabricante tivessem o mesmo valor de resistência, se as determinações de pH de espécimes de solo de um local determinado fornecessem resultados idênticos, e assim por diante, então uma única observação revelaria todas as informações desejadas.

Uma manifestação interessante das variações surge ao longo dos testes de desempenho de emissões em motores automotivos. Os requisitos de custos e tempo do FTP (*Federal Test Procedure*) impedem seu uso generalizado em programas de inspeção veiculares. Como resultado, muitas agências desenvolveram testes mais rápidos e baratos, que, espera-se, reproduzam os resultados do FTP. De acordo com o artigo de jornal "Motor Vehicle Emissions Variability" (*J. of the Air and Waste Mgmt. Assoc.*, 1996: 667-675), a aceitação do FTP como um padrão universal leva à crença de que medidas repetidas no mesmo veículo terão resultados idênticos (ou quase idênticos). Os autores do artigo aplicaram o FTP a sete veículos caracterizados como "altos emissores". Seguem os resultados de um dos veículos:

HC (gm/milha)	13,8	18,3	32,2	32,5
CO (gm/milha)	118	149	232	236

A variação substancial nas medidas de HC e CO apresenta dúvidas consideráveis sobre a sabedoria popular e dificulta a elaboração de avaliações precisas sobre níveis de emissão.

Como as técnicas estatísticas podem ser usadas para obter informações e tirar conclusões? Suponha, por exemplo, que um engenheiro de materiais tenha desenvolvido um revestimento para retardar a corrosão em tubulações de metal sob circunstâncias especificadas. Se esse revestimento for aplicado a diferentes segmentos do tubo, variações nas condições ambientais e nos próprios segmentos resultarão em uma corrosão maior em alguns segmentos do que em outros. Os métodos de análise estatística podem ser usados nos dados de um experimento como esse para decidir se a quantidade média de corrosão excede um limite superior especificado de algum tipo ou para prever a quantidade de corrosão que ocorrerá num único tubo.

Como alternativa, suponha que o engenheiro tenha desenvolvido tal revestimento acreditando que será superior àquele usado no momento. Um experimento comparativo pode ser efetuado para investigar essa questão, aplicando-se o revestimento atual a alguns segmentos do tubo e o novo a outros. Isso deve ser feito com cuidado, para que não surja uma conclusão errada. Por exemplo: talvez a quantidade média de corrosão seja idêntica para os dois revestimentos. Entretanto, o novo revestimento pode ter sido aplicado a segmentos que possuem uma capacidade superior de resistência à corrosão e sob condições ambientais menos severas, se comparados aos segmentos e condições do revestimento atual. O investigador provavelmente observaria então uma diferença causada não pelos próprios revestimentos, mas por variações externas. A estatística oferece métodos não somente para análise dos resultados de experimentos depois que foram executados, como também sugestões de como os experimentos devem ser executados de forma eficiente para diminuir os efeitos das variações e ter melhores chances de produzir conclusões corretas.

1.1 Populações, amostras e processos

Os engenheiros e cientistas estão constantemente expostos a conjuntos de fatos ou **dados**, tanto em suas carreiras como em suas atividades diárias. A disciplina estatística fornece métodos para organizar e resumir os dados para tirar conclusões com base em informações contidas nos dados.

Uma investigação normalmente enfocará uma coleção bem definida de objetos que constituem uma **população** de interesse. Em um estudo, a população pode consistir em todas as cápsulas de gelatina de um determinado tipo produzidas durante um período especificado. Outra investigação pode envolver a população que consiste em todos os indivíduos que receberam um diploma de engenharia durante o ano acadêmico mais recente. Quando as informações desejadas estiverem disponíveis para todos os objetos da população, temos o que é denominado **censo**. Restrições de tempo, dinheiro e outros recursos escassos normalmente tornam um censo impraticável ou inviável. Em vez disso, um subconjunto da população – uma **amostra** – é selecionado de uma forma prescrita. Dessa maneira, podemos obter uma amostra de mancais de uma determinada produção como base de investigação da conformidade dos mancais com as especificações do fabricante; ou podemos selecionar uma amostra dos formandos em engenharia do ano anterior para obter um retorno sobre a qualidade dos currículos.

Normalmente, estamos interessados apenas em certas características dos objetos de uma população: o número de falhas na superfície de cada invólucro, a espessura de cada parede da cápsula, o sexo de um formando em engenharia, a idade com que um indivíduo se formou etc. Uma característica pode ser categorizada, como sexo ou tipo de defeito, ou pode ter natureza numérica. No primeiro caso, o valor da característica é uma categoria (por exemplo, feminino ou solda insuficiente), enquanto, no último caso, o valor é um número (por exemplo, idade = 23 anos ou diâmetro = 0,502 cm). Uma **variável** é qualquer característica cujo valor pode mudar de um objeto para outro na população. Inicialmente, devemos identificar as variáveis com letras minúsculas do final do nosso alfabeto. Os exemplos incluem:

x = marca da calculadora de um estudante
y = número de defeitos graves em um automóvel recentemente fabricado
z = distância de frenagem de um automóvel sob condições especificadas

Os dados resultam da observação de uma variável ou de duas ou mais variáveis simultaneamente. Um conjunto de dados **univariado** consiste em observações sobre uma única variável. Por exemplo: podemos determinar o tipo de transmissão, automática (A) ou manual (M), de cada um dentre 10 automóveis recentemente comprados em um determinado revendedor, resultando em um conjunto de dados categorizados.

M A A A M A A M A A

A amostra a seguir de vida útil (horas) de baterias da marca D colocadas em um determinado uso é um conjunto numérico de dados univariados:

5,6 5,1 6,2 6,0 5,8 6,5 5,8 5,5

Temos dados **bivariados** quando as observações são feitas em cada uma de duas variáveis. Nosso conjunto de dados pode consistir em um par (altura, peso) de cada jogador de basquete de um time, com a primeira observação como (72, 168), a segunda como (75, 212) e assim por diante. Se um engenheiro determinar o valor de x = vida útil do componente e y = motivo de falha do componente, o conjunto de dados resultante será bivariado com uma variável numérica e outra categorizada. Dados **multivariados** surgem quando são feitas observações sobre mais de duas variáveis. Por exemplo: um médico pesquisador pode determinar a pressão sangüínea sistólica, a pressão sangüínea diastólica e o nível de colesterol de cada paciente participante de um estudo. Cada observação seria um trio de números, como (120, 80, 146). Em muitos conjuntos de dados multivariados, algumas variáveis são numéricas e outras são categorizadas. Dessa forma, a edição automotiva anual de *Consumer Reports* fornece valores de tais variáveis como tipo do veículo (pequeno, esportivo, compacto, médio, grande), consumo de combustível na cidade (milhas/galão), consumo de combustível na estrada (milhas/galão), tipo de tração (traseira, dianteira, nas quatro rodas) e assim por diante.

Ramos da estatística

Um investigador que tenha coletado dados pode simplesmente desejar resumir e descrever suas características importantes. Isso exige a utilização de métodos de **estatística descritiva**. Alguns desses métodos são gráficos por natureza: a construção de histogramas, *boxplots* e gráficos de dispersão são os exemplos principais. Outros métodos descritivos envolvem o cálculo de medidas numéricas, como médias, desvios padrão e coeficientes de correlação. A ampla disponibilidade de pacotes de softwares estatísticos facilitou bastante essas tarefas. Os computadores são muito mais eficientes que os seres humanos em cálculo e na criação de imagens (depois de terem recebido as instruções corretas do usuário). Isso significa que o investigador não tem de perder muito tempo com "trabalho braçal" e terá mais tempo para estudar os dados e extrair mensagens importantes. Neste livro serão apresentados resultados de vários pacotes, como MINITAB, SAS, e S-Plus.

Exemplo 1.1

A tragédia que ocorreu com o ônibus espacial *Challenger* e seus astronautas, em 1986, levou a diversos estudos para investigar os motivos da falha da missão. A atenção rapidamente se voltou ao comportamento dos anéis de

vedação do motor do foguete. Aqui estão os dados resultantes de observações de x = temperatura do anel de vedação (°F) de cada teste de acionamento ou lançamento real do motor do foguete da nave (*Presidential Commission on the Space Shuttle Challenger Accident*, Vol. 1, 1986: 129-131).

84	49	61	40	83	67	45	66	70	69	80	58
68	60	67	72	73	70	57	63	70	78	52	67
53	67	75	61	70	81	76	79	75	76	58	31

Sem nenhuma organização, é difícil ter noção do que pode ser uma temperatura normal ou representativa, se os valores estão altamente concentrados em torno de um ponto ou se estão dispersos, se há lacunas nos dados, que porcentagem dos dados estão na faixa dos 60 e assim por diante. A Figura 1.1 mostra o que é chamado de *diagrama caule e folha* dos dados, assim como um *histograma*. Mais adiante serão discutidas a construção e a interpretação desses resumos ilustrativos; no momento, desejamos que você observe como eles começam a nos mostrar como os valores de temperatura são distribuídos ao longo da escala de medida. Alguns desses lançamentos/acionamentos tiveram sucesso e outros resultaram em falha. No Capítulo 13, consideraremos se a temperatura teve influência na probabilidade de um lançamento bem-sucedido.

```
Caule folha da temp N = 36
Unidade da folha = 1,0
 1   3  1
 1   3
 2   4  0
 4   4  59
 6   5  23
 9   5  788
13   6  0113
(7)  6  6777789
16   7  000023
10   7  556689
 4   8  0134
```

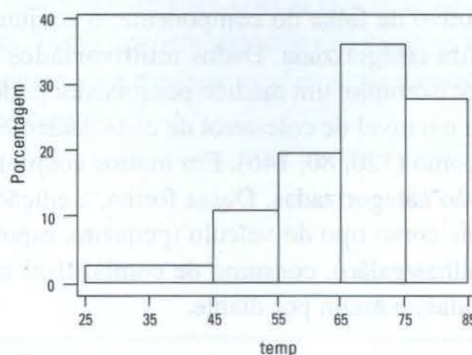


Figura 1.1 Um diagrama caule e folha MINITAB e histograma dos dados da temperatura do anel de vedação ■

Com uma amostra da população, um investigador freqüentemente usaria as informações da amostra para tirar algum tipo de conclusão (fazer uma inferência de algum tipo) sobre a população, ou seja, a amostra é um meio para chegar a um fim e não o fim em si. As técnicas de generalização de uma amostra para uma população são agrupadas no ramo da nossa disciplina denominado **inferência estatística**.

Exemplo 1.2

As investigações de resistência de materiais fornecem um campo fértil para a aplicação de métodos estatísticos. O artigo "Effects of Aggregates and Microfillers on the Flexural Properties of Concrete" (*Magazine of Concrete Research*, 1997, p. 81-98) relatou um estudo de propriedades de resistência de concreto de alto desempenho obtidas pela utilização de superplásticos e determinados adesivos. A resistência à compressão desse concreto foi

investigada anteriormente, mas não se sabe muito sobre a resistência à flexão (uma medida da capacidade de resistência a falhas decorrentes de flexão). Os dados a seguir sobre resistência à flexão (em megapascal, MPa, onde 1 Pa (pascal) = $1{,}45 \times 10^{-4}$ psi) foram exibidos no artigo citado:

5,9	7,2	7,3	6,3	8,1	6,8	7,0	7,6	6,8	6,5	7,0	6,3	7,9	9,0
8,2	8,7	7,8	9,7	7,4	7,7	9,7	7,8	7,7	11,6	11,3	11,8	10,7	

Suponha que busquemos uma *estimativa* do valor médio da resistência à flexão de todas as vigas que podem ser feitas dessa forma (se considerarmos a população de todas as vigas, estaremos tentando estimar a média da população). Pode-se mostrar que, com alto nível de confiança, a resistência média da população está entre 7,48 MPa e 8,80 MPa. Isso é denominado *intervalo de confiança* ou *estimativa por intervalo.* De forma alternativa, esses dados podem ser usados para prever a resistência à flexão de uma *única* viga desse tipo. Com alto nível de confiança, a resistência de uma determinada viga excederá 7,35 MPa. O número 7,35 é denominado *limite inferior de previsão.* ■

Este livro enfoca principalmente os métodos de apresentação e ilustração de inferência estatística úteis ao trabalho científico. Os tipos mais importantes de procedimentos inferenciais (estimativa por pontos, teste de hipóteses e estimativa por intervalos de confiança) são apresentados nos capítulos 6-8 e usados em aplicações mais avançadas nos capítulos 9-16. O restante deste capítulo apresenta os métodos de estatística descritiva mais usados no desenvolvimento da inferência.

Os capítulos 2-5 apresentam material da disciplina de probabilidade. Esse material, em suma, faz uma ponte entre as técnicas descritiva e inferencial. A proficiência em probabilidade leva à melhor compreensão de como os procedimentos inferenciais são desenvolvidos e usados, como as conclusões estatísticas podem ser traduzidas para a linguagem do dia-a-dia e interpretadas, e quando e onde podem ocorrer ciladas na aplicação dos métodos. A probabilidade e a estatística lidam com questões que envolvem populações e amostras, mas o fazem de "maneira inversa" uma em relação a outra.

Em um problema de probabilidade, as propriedades da população, que são objeto de estudo, são assumidas como conhecidas (por exemplo: em uma população numérica, uma distribuição especificada dos valores da população pode ser assumida) e as questões relativas a uma amostra proveniente da população são propostas e respondidas. Em um problema de estatística, as características de uma amostra estão disponíveis ao investigador e essas informações permitem que ele tire conclusões sobre a população. A relação entre as duas disciplinas pode ser resumida da seguinte forma: a probabilidade faz suas considerações da população para a amostra (raciocínio dedutivo) e a inferência estatística faz considerações da amostra para a população (raciocínio indutivo). Isso é ilustrado na Figura 1.2.

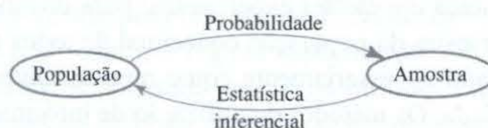


Figura 1.2 A relação entre probabilidade e inferência estatística

Antes de podermos entender o que uma determinada amostra pode nos dizer sobre a população, devemos entender a incerteza associada à tomada da amostra de uma dada população. É por isso que estudamos probabilidade antes de estatística.

Como exemplo do contraste entre os focos da probabilidade e da inferência estatística, considere o uso de cintos de segurança manuais de dois pontos em carros equipados com cintos automáticos de três pontos. (O artigo "Automobile Seat Belts: Usage Patterns in Automatic Belt Systems," *Human Factors,* 1998, p. 126-135, resume os dados de utilização.) Em probabilidade, podemos assumir que 50% de todos os motoristas de carros equipados dessa forma em uma determinada área metropolitana usam regularmente o cinto de dois pontos (uma hipótese sobre a população), de forma que perguntamos: "Qual a probabilidade de que uma amostra de 100 motoristas inclua ao menos 70 que usam regularmente o cinto de dois pontos?" ou "Em uma amostra de tamanho

100, quantos motoristas podemos esperar que usem o cinto de dois pontos?" Por outro lado, em inferência estatística temos as informações da amostra disponíveis. Por exemplo: uma amostra de 100 motoristas de tais carros revelou que 65 usam o cinto de dois pontos regularmente. Podemos perguntar então: "Isso fornece evidência suficiente para a conclusão de que mais de 50% de todos os motoristas nessa área usam regularmente o cinto de dois pontos?" Nesse último cenário, tentamos usar as informações da amostra para responder a uma pergunta sobre a estrutura de toda a população a partir da qual a amostra foi selecionada.

No exemplo do cinto de dois pontos, a população está bem definida e concreta: todos os motoristas de carros equipados de uma forma em uma determinada área metropolitana. No Exemplo 1.1, entretanto, está disponível uma amostra de temperaturas de anéis de vedação de uma população que não existe realmente. Em vez disso, é conveniente pensarmos na população como consistindo de todas as medidas possíveis de temperatura que podem ser feitas em condições experimentais similares. Tal população é denominada **população conceitual** ou **hipotética**. Há diversas situações de problemas em que se encaixam questões na estrutura de inferência estatística pela conceitualização de uma população.

Estudos enumerativos *versus* analíticos

W. E. Deming, um influente estatístico norte-americano, força motriz na revolução de qualidade do Japão nos anos 50 e 60, apresentou a distinção entre *estudos enumerativos* e *estudos analíticos*. No primeiro, o interesse se foca em uma coleção finita, identificável e imutável de indivíduos ou objetos que formam uma população. Uma *estrutura de amostragem*, isto é, uma listagem de todos os indivíduos ou objetos a fazerem parte da amostra, está disponível a um investigador ou pode ser construída. Por exemplo: a estrutura pode ser constituída por todas as assinaturas em uma petição para qualificação de uma determinada iniciativa de voto secreto em uma eleição iminente; uma amostra normalmente é selecionada para apurar se o número de assinaturas *válidas* excede um valor especificado. Como outro exemplo, a estrutura pode conter números de série de todos os fornos fabricados por uma determinada empresa durante certo período de tempo; uma amostra pode ser selecionada para inferir algo sobre a vida útil média dessas unidades. A utilização de métodos inferenciais a ser desenvolvida neste livro é razoavelmente não-controversa nesse cenário (apesar de os estatísticos ainda poderem discutir sobre que métodos em particular devem ser usados).

Um estudo analítico é definido, de modo geral, como aquele que não é de natureza enumerativa. Esses estudos são normalmente executados com o objetivo de melhorar um produto futuro por meio da ação em um processo de algum tipo (por exemplo: recalibragem de equipamentos ou ajuste do nível de algum insumo, como a quantidade de um catalisador). Freqüentemente, os dados podem ser obtidos apenas em um processo existente, que pode diferir do processo futuro em aspectos importantes. Dessa forma, não há uma estrutura de amostragem que relacione os indivíduos ou objetos de interesse. Por exemplo: uma amostra de cinco turbinas com um novo *design* pode ser fabricada e testada em caráter experimental para investigar sua eficiência. Essas cinco turbinas podem ser vistas como uma amostra da população conceitual de todos os protótipos que podem ser fabricados em condições similares, mas *não* necessariamente como representantes da população de unidades fabricadas depois que a produção for iniciada. Os métodos de utilização de informações de amostras para obtenção de conclusões sobre a produção futura podem ser problemáticos. Alguém com experiência na área de projetos e engenharia de turbinas (ou de qualquer outra área relevante para a disciplina) deve ser chamado para julgar se essa extrapolação é sensata. Uma boa exposição dessas questões está no artigo "Assumptions for Statistical Inference" de Gerald Hahn e William Meeker (*The American Statistician*, 1993, p. 1-11).

Coletando dados

A estatística lida não somente com a organização e análise de dados depois de sua coleta, como também com o desenvolvimento de técnicas de coleta. Se os dados não são coletados de forma correta, um investigador pode não ter condições de responder às questões em consideração com um nível de confiança razoável. Um problema comum é que a população-alvo — a respeito da qual serão tiradas conclusões — pode ser diferente da população da qual se obteve a amostra. Por exemplo: publicitários podem desejar diversos tipos de informações sobre os hábitos televisivos de clientes potenciais. As informações mais sistemáticas desse tipo são provenientes de dispositivos de monitoramento locais em um pequeno número de lares nos Estados Unidos. Já se presumiu que a

colocação em si desses dispositivos afeta o comportamento dos telespectadores, de forma que as características da amostra podem ser diferentes daquelas da população-alvo.

Quando a coleta de dados exige a seleção de indivíduos ou objetos a partir de uma estrutura, o método mais simples de assegurar uma seleção representativa é tomar uma *amostra aleatória simples*. Trata-se de uma amostra em que qualquer subconjunto de tamanho especificado (como uma amostra de tamanho 100) tem a mesma chance de ser selecionado. Por exemplo: se a estrutura consistir em 1.000.000 de números de série, os números 1, 2, ... até 1.000.000 podem ser colocados em tiras idênticas de papel. Após colocá-las em uma caixa e mexer bastante, elas podem ser retiradas uma a uma, até que a amostra de tamanho requerido seja obtida. De forma alternativa (e normalmente preferida), uma tabela de números aleatórios ou um gerador de números aleatórios pode ser usado.

Algumas vezes, os métodos de amostragem alternativos podem ser usados para facilitar o processo de seleção, para obter informações extras ou para aumentar o nível de confiança das conclusões. Um desses métodos, a *amostragem estratificada*, exige a separação das unidades da população em grupos não-passíveis de sobreposição e a tomada de uma amostra de cada um. Por exemplo: um fabricante de videocassetes pode desejar informações sobre a satisfação dos clientes com as unidades produzidas no ano anterior. Se foram fabricados e vendidos três modelos diferentes, pode ser selecionada uma amostra de cada um dos três modelos correspondentes, o que resultaria em informações sobre todos os modelos e asseguraria que nenhum deles teve mais ou menos representatividade na amostra inteira.

Freqüentemente, uma amostra de "conveniência" é obtida pela seleção de indivíduos ou objetos sem aleatoriedade sistemática. Como exemplo, um grupo de tijolos pode ser empilhado de forma que seja extremamente difícil selecionar as peças centrais. Se os tijolos do topo e das laterais forem de alguma forma diferentes dos outros, os dados resultantes da amostra não serão representativos da população. Um investigador assumirá, com freqüência, que essa amostra de conveniência se aproxima de uma amostra aleatória. Nesse caso, o repertório de métodos inferenciais do estatístico pode ser usado, o que é, entretanto, de julgamento do profissional. A maioria dos métodos discutidos daqui em diante são baseadas em variações da amostragem aleatória simples, descritas no Capítulo 5.

Os engenheiros e cientistas freqüentemente coletam dados executando algum tipo de experimento, o que pode envolver a decisão de como alocar diferentes tratamentos (como fertilizantes ou revestimentos para proteção de corrosão) às diversas unidades experimentais (lotes de terra ou segmentos de tubo). Além disso, um investigador pode variar sistematicamente os níveis ou categorias de determinados fatores (como pressão ou tipo de material isolante) e observar o efeito em alguma variável de resposta (como o resultado de um processo de produção).

Exemplo 1.3

Um artigo no *New York Times* (27 de janeiro de 1987) reportou que o risco de ataques cardíacos pode ser reduzido pela ingestão de aspirina. Essa conclusão foi baseada em um experimento planejado que envolveu um grupo de controle de indivíduos que tomaram um placebo com aparência de aspirina, mas reconhecidamente inócuo, e outro que tomou aspirina de acordo com um regime especificado. Os comprimidos foram atribuídos aleatoriamente aos grupos para evitar desvios, tornando possível a utilização de métodos probabilísticos para análise dos dados. Dos 11.034 componentes do grupo de controle, 189 tiveram ataques cardíacos posteriormente, enquanto apenas 104 dos 11.037 indivíduos do grupo da aspirina tiveram o problema. A taxa de incidência de ataques cardíacos no grupo de tratamento foi cerca de metade da taxa do grupo de controle. Uma explicação possível para esse resultado é a variação devida ao acaso: que a aspirina não tem o efeito desejado e que a diferença observada é uma variação normal, da mesma forma que jogar duas moedas idênticas normalmente produziria números diferentes de caras. Entretanto, nesse caso, os métodos inferenciais sugerem que a variação devida ao acaso em si não pode explicar adequadamente a magnitude da diferença observada. ■

Exemplo 1.4

Um engenheiro deseja investigar os efeitos de um tipo de adesivo e de um material condutor na resistência do contato ao montar um IC (Circuito Integrado) em uma determinada base. Dois tipos de adesivo e dois materiais condutores estão sendo considerados. Duas observações são feitas para cada combinação – tipo de adesivo/material condutor – resultando nos dados a seguir:

Tipo de adesivo	Material condutor	Resistência observada do contato	Média
1	1	82, 77	79,5
1	2	75, 87	81,0
2	1	84, 80	82,0
2	2	78, 90	84,0

As resistências de contato médias são ilustradas na Figura 1.3. Parece que o tipo de adesivo 2 melhora a resistência do contato se comparado ao tipo 1 mais ou menos no mesmo valor, não importando o material condutor usado, com a combinação 2, 2 sendo a melhor. Os métodos inferenciais podem novamente ser usados para julgar se esses efeitos são reais ou simplesmente conseqüência da variância devido ao acaso.

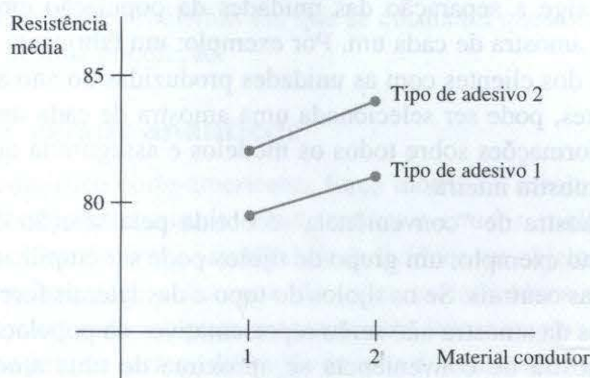


Figura 1.3 Resistências de contato médias no Exemplo 1.4

Suponha, adicionalmente, que haja dois períodos de cura em consideração e também dois tipos de circuito integrado após o revestimento. Há, portanto, $2 \cdot 2 \cdot 2 \cdot 2 = 16$ combinações desses quatro fatores e nosso engenheiro não possui recursos suficientes nem mesmo para fazer uma única observação para cada uma dessas combinações. No Capítulo 11 veremos como a seleção cuidadosa de uma fração dessas possibilidades normalmente fornecerá as informações desejadas. ■

Exercícios | Seção 1.1 (1–9)

1. Forneça uma amostra possível, de tamanho 4, de cada uma das populações a seguir:

a. Todos os jornais diários publicados nos Estados Unidos.

b. Todas as empresas listadas na New York Stock Exchange.

c. Todos os alunos de sua universidade ou faculdade.

d. Todas as médias, em pontos, dos alunos de sua universidade ou faculdade.

2. Para cada uma das populações hipotéticas a seguir, forneça uma amostra plausível de tamanho 4:

a. Todas as distâncias que podem resultar quando uma bola de futebol é arremessada.

b. O tamanho das páginas dos livros publicados nos últimos cinco anos.

c. Todas as medidas de intensidade de terremotos (escala Richter) que podem ser registradas na Califórnia durante o próximo ano.

d. Todos os possíveis produtos (em gramas) de uma determinada reação química feita em um laboratório.

3. Considere a população que consiste em todos os videocassetes de uma determinada marca e modelo, e enfoque se um videocassete precisa de manutenção durante o tempo de garantia.

a. Formule diversas questões sobre probabilidade baseadas em uma amostra de 100 desses videocassetes.

b. Qual questão sobre inferência estatística pode ser respondida ao determinar o número de videocassetes que precisam de serviço de garantia em uma amostra de tamanho 100?

4. a. Dê três exemplos diferentes de populações concretas e três exemplos diferentes de populações hipotéticas.

b. Para cada uma de suas populações hipotéticas e concretas, dê um exemplo de uma questão sobre probabilidade e um exemplo de uma questão sobre inferência estatística.

5. Diversas universidades e faculdades instituíram programas de Instrução Suplementar (SI), em que um monitor se encontra regularmente com um grupo de estudantes matriculados em um curso para promover discussões

sobre o material desse curso e melhorar o domínio da disciplina. Suponha que os estudantes de um grande curso de estatística (o que mais poderia ser?) são aleatoriamente divididos em um grupo de controle que não participará do SI e um grupo de tratamento que o fará. No final do período, é determinada a pontuação total de cada estudante no curso.

a. As pontuações do grupo de SI são uma amostra da população existente? Caso seja, qual é? Caso contrário, qual é a população conceitual relevante?

b. Qual você acha que é a vantagem de dividir aleatoriamente os estudantes em dois grupos em vez de deixar cada estudante escolher o grupo do qual participará?

c. Por que os investigadores não colocaram todos os estudantes no grupo de tratamento? *Nota:* O artigo "Supplemental Instruction: An Effective Component of Student Affairs Programming" (*J. of College Student Devel.*, 1997, p. 577-586) discute a análise de dados de diversos programas de SI.

6. O sistema da CSU (California State University) consiste em 23 *campi*, de San Diego State, no sul, até Humboldt State, perto da fronteira com Oregon. Um administrador da CSU deseja fazer uma inferência sobre a distância média entre as cidades natais de seus alunos e seus *campi*. Descreva e discuta diversos métodos de amostragem que podem ser empregados. Esse estudo seria enumerativo ou analítico? Explique seu raciocínio.

7. Certa cidade é dividida naturalmente em 10 bairros. Como um avaliador imobiliário deve selecionar uma amostra de casas de uma única família que pode ser usada como base para o desenvolvimento de uma equação para previsão do valor avaliado a partir de características como idade, tamanho, número de banheiros, distância até a escola mais próxima e assim por diante? Esse estudo é enumerativo ou analítico?

8. A quantidade de fluxo que passa através de uma válvula solenóide em um sistema de controle de poluição de um automóvel é uma característica importante. Foi executado um experimento para estudar como a taxa do fluxo depende de três fatores: comprimento do núcleo, carga da mola e largura da bobina. Foram escolhidos dois níveis diferentes (alto e baixo) de cada fator e foi feita uma única observação sobre o fluxo para cada combinação de níveis.

a. O conjunto de dados resultante consistiu em quantas observações?

b. Esse estudo é enumerativo ou analítico? Explique seu raciocínio.

9. Em um famoso experimento executado em 1882, Michelson e Newcomb fizeram 66 observações do tempo levado pela luz para percorrer a distância entre dois locais em Washington, D.C. Algumas das medidas (codificadas de certa forma) foram 31, 23, 32, 36, –2, 26, 27 e 31.

a. Por que essas medidas não são idênticas?

b. Esse estudo é enumerativo? Por quê?

1.2 Métodos tabular e gráfico em estatística descritiva

A estatística descritiva pode ser dividida em duas áreas gerais. Nesta seção, discutiremos a primeira dessas áreas, que representa um conjunto de dados usando técnicas visuais. Nas seções 1.3 e 1.4, desenvolveremos algumas medidas numéricas simples para conjuntos de dados. Muitas técnicas visuais podem ser conhecidas: tabelas de freqüência, folhas de contagem, histogramas, gráficos de pizza, gráficos de barras, digramas de dispersão e afins. Aqui enfocamos algumas dessas técnicas, que são mais úteis e relevantes para probabilidade e inferência estatística.

Notações

Algumas notações gerais facilitarão a aplicação de nossos métodos e fórmulas a uma ampla gama de problemas práticos. O número de observações em uma única amostra, isto é, o *tamanho da amostra*, normalmente será representado por n, de forma que $n = 4$ para a amostra de universidades {Stanford, Iowa State, Wyoming, Rochester} e também para a amostra de medidas de pH {6,3; 6,2; 5,9; 6,5}. Se duas amostras estiverem sendo consideradas simultaneamente, m e n ou n_1 e n_2 podem ser usados para representar os números de observações. Portanto, se {29,7; 31,6; 30,9} e {28,7; 29,5; 29,4; 30,3} forem medidas de eficiência térmica para dois tipos de motor diesel, então $m = 3$ e $n = 4$.

Dado um conjunto de dados que consiste de n observações de uma variável x, as observações individuais serão representadas por $x_1, x_2, x_3, \ldots, x_n$. O índice não tem nenhuma relação com a magnitude de uma

determinada observação. Dessa forma, x_1 em geral não será a menor observação do conjunto e x_n normalmente não será a maior. Em diversas aplicações, x_1 será a primeira observação coletada pelo investigador, x_2 será a segunda e assim por diante. A *i*-ésima observação do conjunto de dados será representada por x_i.

Diagramas de caule e folha

Considere um conjunto de dados numéricos $x_1, x_2, \dots, x_n$, para o qual cada x_i consiste de, pelo menos, dois dígitos. Uma forma rápida de obter uma representação visual informativa do conjunto de dados é construir um *diagrama de caule e folha.*

Etapas de construção de um diagrama de caule e folha

1. Selecione um ou mais dígitos de liderança para serem o caule. Os dígitos à direita desse dígito de liderança serão as folhas.
2. Relacione os valores de caule possíveis em uma coluna vertical.
3. Registre a folha de cada observação ao lado do caule correspondente.
4. Indique as unidades dos caules e das folhas em algum lugar do diagrama.

Se o conjunto de dados consistir em notas de provas, cada uma entre 0 e 100, a pontuação de 83 terá caule 8 e folha 3. Para um conjunto de dados de consumo de combustível de automóveis (milhas/galão), todos entre 8,1 e 47,8, podemos usar as dezenas como caule, de forma que 32,6 teria uma folha de 2,6. Em geral, recomenda-se que o diagrama tenha entre 5 e 20 caules.

Exemplo 1.5

O consumo de álcool por alunos de faculdades causa grande preocupação, não apenas para os membros da comunidade acadêmica, como também pelas conseqüências potenciais à saúde e à segurança da sociedade em geral. O artigo "Health and Behavioral Consequences of Binge Drinking in College" (*J. of the Amer Med. Assoc.*, 1994, p. 1672-1677) relatou em um abrangente estudo de consumo excessivo de álcool em diversos *campi* nos Estados Unidos. Um episódio de bebedeira foi definido como cinco ou mais bebidas em seqüência para os homens e quatro ou mais para as mulheres. A Figura 1.4 mostra um diagrama de caule e folha de 140 valores de $x =$ ao percentual de estudantes de nível superior que se embriagam. (Esses valores não foram fornecidos no artigo citado, mas nosso diagrama apresenta-se de acordo com a ilustração exibida dos dados.)

A primeira folha na linha do caule 2 é 1, o que nos informa que 21% dos alunos de uma das faculdades da amostra se embriagavam. Sem a identificação dos dígitos do caule e das folhas no diagrama, não saberíamos se a observação do caule 2, folha 1, devia ser lida como 21%, 2,1% ou 0,21%.

0	4
1	1345678889
2	1223456666777889999
3	0112233344555666677777888899999
4	111222223344445566666677788888999
5	00111222233455666667777888899
6	01111244455666778

Caule: dígito das dezenas
Folha: dígito das unidades

Figura 1.4 Diagrama de caule e folha de percentual de alunos que se embriagam em cada uma de 140 faculdades

Ao criar um diagrama a mão, a organização das folhas da menor para a maior em cada linha pode tomar muito tempo e essa organização normalmente contribui pouco ou quase nada para informações extras. Suponha que as observações tenham sido relacionadas em ordem alfabética por nome de escola, da seguinte forma:

16% 33% 64% 37% 31% . . .

Colocar esses valores no diagrama nessa ordem resultaria no caule 1 tendo 6 como sua primeira folha e o começo da linha do caule 3 seria

3 | 371 . . .

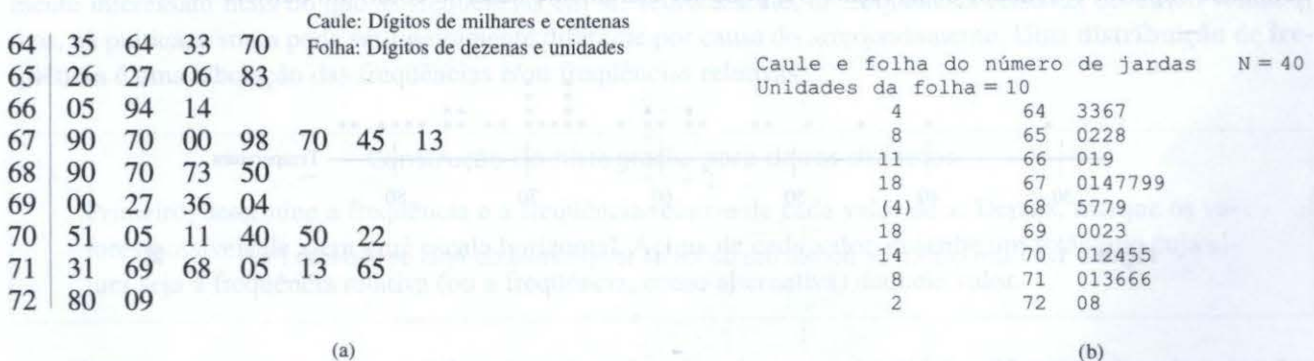
O diagrama sugere que um valor típico ou representativo está na linha do caule 4, talvez na metade da faixa de 40%. As observações não estão altamente concentradas em torno desse valor típico, como estariam se todos os valores estivessem entre 20% e 49%. O diagrama cresce para um único pico, quando nos movemos para baixo e então declina: não há lacunas no diagrama. Seu formato não é perfeitamente simétrico, parece aumentar mais na direção das folhas inferiores, do que na direção das folhas de cima. Por último, não há observações distantes da parte principal dos dados (sem *outliers*, ou seja, pontos fora da curva), como aconteceria se um dos valores fosse 86% em vez de 26%. A característica mais surpreendente dos dados é que, na maior parte das faculdades, pelo menos um quarto dos alunos se embriaga. O problema do consumo excessivo de álcool nos *campi* é muito mais difundido do que muitos haviam suspeitado. ■

Um diagrama de caule e folha transmite informações sobre os seguintes aspectos dos dados:

- identificação de um valor típico ou representativo;
- extensão da dispersão ao redor do valor típico;
- presença de lacunas nos dados;
- extensão da simetria na distribuição de valores;
- número e localização dos bicos;
- presença de valores fora da curva.

Exemplo 1.6

A Figura 1.5 apresenta diagramas de caule e folha para uma amostra aleatória de comprimentos de campos de golfe (jardas) que foram designados pela *Golf Magazine* entre os mais desafiadores dos Estados Unidos. Entre a amostra de 40 campos, o mais curto tem 6.433 jardas de comprimento e o mais longo, 7.280. Os comprimentos parecem estar distribuídos de forma aproximadamente uniforme acima da faixa de valores da amostra. Observe que, aqui, uma escolha de caule de um único dígito (6 ou 7) ou de três dígitos (643, ... , 728) resultaria em diagramas não-informativos, no primeiro caso porque não haveria caules suficientes, no segundo, porque haveria caules demais.



Caule: Dígitos de milhares e centenas
Folha: Dígitos de dezenas e unidades

64	35	64	33	70			
65	26	27	06	83			
66	05	94	14				
67	90	70	00	98	70	45	13
68	90	70	73	50			
69	00	27	36	04			
70	51	05	11	40	50	22	
71	31	69	68	05	13	65	
72	80	09					

(a)

```
Caule e folha do número de jardas   N = 40
Unidades da folha = 10
     4      64  3367
     8      65  0228
    11      66  019
    18      67  0147799
   (4)      68  5779
    18      69  0023
    14      70  012455
     8      71  013666
     2      72  08
```

(b)

Figura 1.5 Diagramas de caule e folha de comprimentos em jardas de campos de golfe: (a) folhas de dois dígitos; (b) diagrama de MINITAB com folhas truncadas de um dígito ■

Um diagrama de caule e folha não mostra a ordem em que as observações foram obtidas, possivelmente ocultando importantes informações sobre o mecanismo gerador dos dados. Por exemplo: durante um período de tempo determinado, a largura de um calço de certa peça feito por uma fresa tende a aumentar em relação ao valor-alvo, indicando um processo "fora de controle". Um *gráfico de linha com marcadores* combina o quadro das observações ao longo do tempo com um diagrama caule e folha.

Exemplo 1.7

Cada observação no diagrama de caule e folha da Figura 1.6 é o valor de produção de cerveja dos EUA (milhões de barris) para um trimestre diferente durante o período de 1975-1982. O diagrama usa *caules repetidos*; por exemplo: a linha 4L é para observações com uma folha "baixa" — 0, 1, 2, 3 ou 4 — e as observações com folhas mais altas são colocadas na linha 4H. O *gráfico da série temporal* à direita mostra tendência de aumento com o tempo e também maior produção nos segundo e terceiro trimestres de um determinado ano do que nos outros dois trimestres (um efeito *sazonal*).

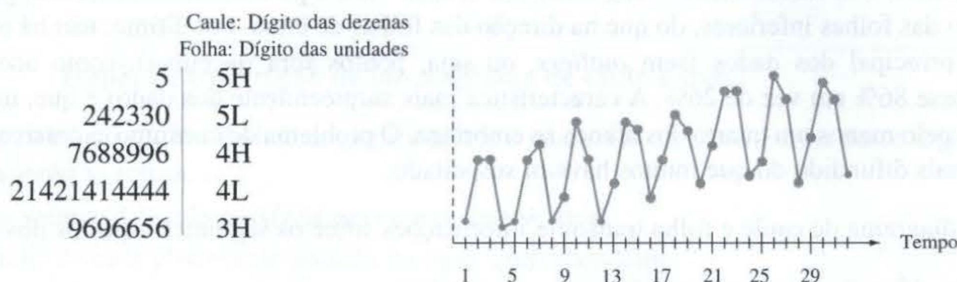


Figura 1.6 Um gráfico de linha com marcadores de produção de cerveja dos Estados Unidos ■

Gráfico de pontos

Um gráfico de pontos é um interessante resumo dos dados numéricos quando esse conjunto é razoavelmente pequeno ou possui relativamente poucos valores distintos. Cada observação é representada por um ponto sobre o local correspondente em uma escala de medida horizontal. Quando um valor ocorre mais de uma vez, há um ponto para cada ocorrência e esses pontos são empilhados verticalmente. Como ocorre com o diagrama de caule e folha, um gráfico de pontos fornece informações sobre localização, dispersão, extremos e lacunas.

Exemplo 1.8

A Figura 1.7 mostra um gráfico de pontos para os dados de temperatura de Anel de Vedação apresentados no Exemplo 1.1, na seção anterior. Um valor de temperatura representativo é o que se apresenta na metade de 60 (°F) e há grande dispersão ao redor do centro. Os dados se estendem mais na extremidade inferior do que na extremidade superior e a menor observação, 31, pode ser claramente descrita como um *outlier*.

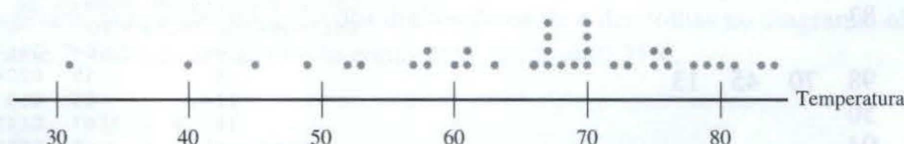


Figura 1.7 Um gráfico de pontos dos dados de temperatura do Anel de Vedação (°F) ■

Se o conjunto de dados discutido no Exemplo 1.8 consistisse de 50 ou 100 observações de temperatura, cada uma registrada em décimos de graus, seria muito mais trabalhoso construir um gráfico de pontos. Nossa próxima técnica é melhor adaptável a tais situações.

Histogramas

Alguns dados numéricos são obtidos por contagem para determinar o valor de uma variável (o número de autuações de tráfego que uma pessoa recebeu no último ano, o número de pessoas que chegam para trabalhar durante um determinado período), enquanto outros dados são obtidos pela tomada de medidas (peso de um indivíduo,

tempo de reação a determinado estímulo). A recomendação para plotagem de um histograma geralmente é diferente para esses dois casos.

DEFINIÇÃO

Uma variável é **discreta** se o seu conjunto de valores possíveis é finito ou pode ser relacionado em uma seqüência infinita (em que haja um primeiro número, um segundo e assim por diante). Uma variável é **contínua** se os seus valores possíveis consistem de um intervalo completo na reta de numeração.

Uma variável discreta x quase sempre é resultante de contagem e, nesse caso, 0, 1, 2, 3... ou algum subconjunto desses inteiros são valores possíveis. Variáveis contínuas surgem da tomada de medidas. Por exemplo: se x é o pH de uma substância química, em teoria, x pode ser qualquer número entre 0 e 14: 7.0, 7.03, 7.032 e assim por diante. Claro que, na prática, há limitações no que diz respeito ao grau de precisão de qualquer instrumento de medida, de forma que podemos não ser capazes de determinar pH, tempo de reação, altura e concentração com um número arbitrariamente grande de casas decimais. Entretanto, do ponto de vista da criação de modelos matemáticos para a distribuição dos dados, é útil imaginar um intervalo contínuo de valores possíveis.

Considere os dados constituídos de observações de uma variável discreta x. A **freqüência** de qualquer valor particular de x é o número de vezes em que esse valor ocorre naquele conjunto. A **freqüência relativa** de um valor é a fração ou proporção de vezes em que o valor ocorre:

$$\text{freqüência relativa de um valor} = \frac{\text{número de vezes que o valor ocorre}}{\text{número de observações do conjunto de dados}}$$

Suponha, por exemplo, que o nosso conjunto de dados consista em 200 observações de x = o número de defeitos graves em um novo carro de certo tipo. Se 70 desses valores x forem 1, então

freqüência do valor $x = 1$: 70

freqüência relativa do valor $x = 1$: $\frac{70}{200} = 0{,}35$

A multiplicação da freqüência relativa por 100 fornece a porcentagem. No exemplo dos defeitos, 35% dos carros da amostra apresentaram apenas um defeito grave. As freqüências relativas, ou porcentagens, normalmente interessam mais do que as freqüências em si. Teoricamente, as freqüências relativas deveriam somar 1, mas, na prática, a soma pode ser ligeiramente diferente por causa do arredondamento. Uma **distribuição de freqüência** é uma tabulação das freqüências e/ou freqüências relativas.

Construção do histograma para dados discretos

Primeiro, determine a freqüência e a freqüência relativa de cada valor de x. Depois, marque os valores possíveis de x em uma escala horizontal. Acima de cada valor, desenhe um retângulo cuja altura seja a freqüência relativa (ou a freqüência, como alternativa) daquele valor.

Essa construção assegura que a *área* de cada retângulo seja proporcional à freqüência relativa do valor. Assim, se as freqüências relativas de $x = 1$ e $x = 5$ são 0,35 e 0,07, respectivamente, a área do retângulo acima de 1 será cinco vezes a área do retângulo acima de 5.

Exemplo 1.9

Quão incomum é um jogador que não atinge a bola ou a atinge uma única vez em um jogo de beisebol da liga principal e com que freqüência uma equipe consegue atingir a bola mais de 10, 15 ou mesmo 20 vezes? A Tabela 1.1 é uma distribuição de freqüência do número de acertos por equipe, por partida, para todos os jogos de nove séries entre 1989 e 1993.

Tabela 1.1 Distribuição de freqüência de acertos em jogos de nove séries

Acertos/Jogo	Número de Jogos	Freqüência Relativa	Acertos/Jogo	Número de Jogos	Freqüência Relativa
0	20	0,0010	14	569	0,0294
1	72	0,0037	15	393	0,0203
2	209	0,0108	16	253	0,0131
3	527	0,0272	17	171	0,0088
4	1048	0,0541	18	97	0,0050
5	1457	0,0752	19	53	0,0027
6	1988	0,1026	20	31	0,0016
7	2256	0,1164	21	19	0,0010
8	2403	0,1240	22	13	0,0007
9	2256	0,1164	23	5	0,0003
10	1967	0,1015	24	1	0,0001
11	1509	0,0779	25	0	0,0000
12	1230	0,0635	26	1	0,0001
13	834	0,0430	27	1	0,0001
				19,383	1,0005

O histograma correspondente da Figura 1.8 tem um leve aclive para um único pico e depois tem um declive. O histograma se estende um pouco mais do lado direito (em direção aos valores maiores) do que para o lado esquerdo – uma inclinação ligeiramente "positiva".

A partir das informações tabuladas ou do histograma em si, podemos determinar o seguinte:

$$\text{proporção de jogos com no máximo dois acertos} = \begin{matrix}\text{freqüência}\\ \text{relativa para}\\ x=0\end{matrix} + \begin{matrix}\text{freqüência}\\ \text{relativa para}\\ x=1\end{matrix} + \begin{matrix}\text{freqüência}\\ \text{relativa para}\\ x=2\end{matrix}$$

$$= 0{,}0010 + 0{,}0037 + 0{,}0108 = 0{,}0155$$

De forma similar,

$$\text{proporção de jogos com acertos entre 5 e 10 (inclusive)} = 0{,}0752 + 0{,}1026 + \ldots + 0{,}1015 = 0{,}6361$$

Isto é, cerca de 64% de todos esses jogos tiveram entre 5 e 10 (inclusive) acertos.

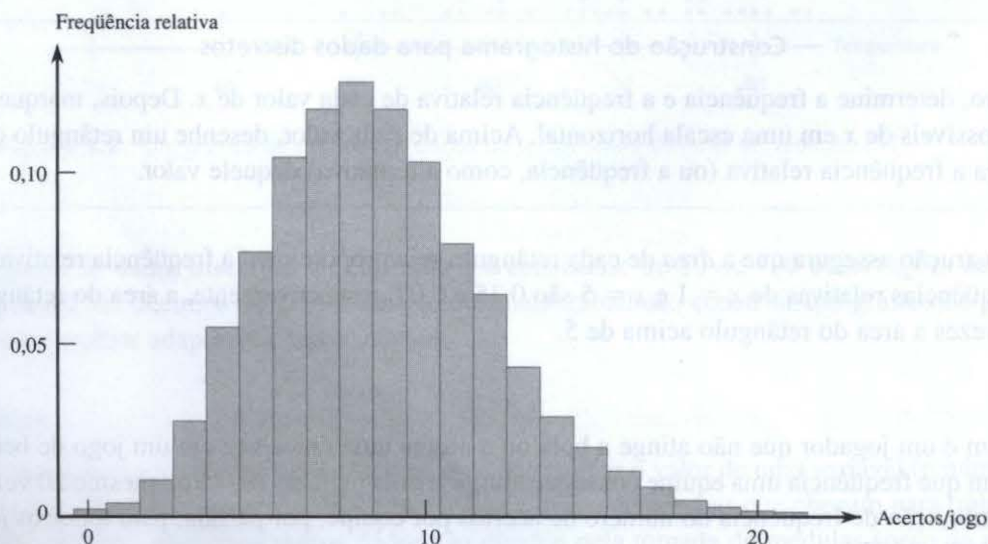
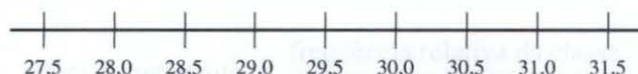


Figura 1.8 Histograma do número de acertos por jogo de nove séries ■

A construção de um histograma de dados contínuos (medidas) exige que o eixo das medidas seja subdividido em um número aceitável de **intervalos de classe** ou **classes**, de forma que cada observação esteja contida completamente em uma classe. Suponha, por exemplo, que tenhamos 50 observações de x = consumo de combustível de um automóvel (milhas/galão), sendo o menor deles 27,8 e o maior, 31,4. Podemos, então, usar os limites de classe 27,5, 28,0, 28,5, ..., e 31,5, conforme mostrado abaixo:



Uma dificuldade potencial é que, ocasionalmente, uma observação fica exatamente sobre um dos limites de classe, não estando necessariamente em apenas um intervalo, por exemplo, 29,0. Uma forma de lidar com esse problema é usar limites como 27,55, 28,05, ... , 31,55. A adição do dígito de centésimos aos limites de classe evita que as observações estejam exatamente sobre os limites resultantes. Outra abordagem é usar as classes 27,5–<28,0, 28,0–<28,5, ... , 31,0–<31,5. Então, 29,0 estaria na classe 29,0–<29,5 em vez de na classe 28,5–< 29,0. Em outras palavras, com essa convenção, uma observação sobre um limite é colocada no intervalo à direita do limite. Essa é a maneira pela qual o MINITAB constrói um histograma.

Construção de histograma para dados contínuos: classes de larguras iguais

Determine a freqüência e a freqüência relativa de cada classe. Marque os limites de classe em um eixo de medida horizontal. Acima de cada intervalo de classe, desenhe um retângulo cuja altura seja a freqüência relativa correspondente (ou a freqüência).

Exemplo 1.10

As empresas de energia necessitam de informações sobre o consumo de seus clientes para obterem previsões precisas da demanda. Investigadores da Wisconsin Power and Light determinaram que o consumo de energia (BTUs) dura um determinado período para uma amostra de 90 lares aquecidos a gás. O valor de consumo ajustado foi calculado conforme segue:

$$\text{consumo ajustado} = \frac{\text{consumo}}{\text{(clima, em grau dias)(área da casa)}}$$

O resultado apresenta-se nos dados anexos (parte do conjunto de dados armazenado FURNACE.MTW disponível no MINITAB) que ordenamos do menor para o maior.

2,97	4,00	5,20	5,56	5,94	5,98	6,35	6,62	6,72	6,78
6,80	6,85	6,94	7,15	7,16	7,23	7,29	7,62	7,62	7,69
7,73	7,87	7,93	8,00	8,26	8,29	8,37	8,47	8,54	8,58
8,61	8,67	8,69	8,81	9,07	9,27	9,37	9,43	9,52	9,58
9,60	9,76	9,82	9,83	9,83	9,84	9,96	10,04	10,21	10,28
10,28	10,30	10,35	10,36	10,40	10,49	10,50	10,64	10,95	11,09
11,12	11,21	11,29	11,43	11,62	11,70	11,70	12,16	12,19	12,28
12,31	12,62	12,69	12,71	12,91	12,92	13,11	13,38	13,42	13,43
13,47	13,60	13,96	14,24	14,35	15,12	15,24	16,06	16,90	18,26

Deixamos o MINITAB selecionar os intervalos de classe. A característica mais impressionante do histograma da Figura 1.9 é sua semelhança com uma curva em forma de sino (e portanto simétrica), com o ponto de simetria próximo a 10.

Classe	1–<3	3–<5	5–<7	7–<9	9–<11	11–<13	13–<15	15–<17	17–<19
Freqüência	1	1	11	21	25	17	9	4	1
Freqüência relativa	0,011	0,011	0,122	0,233	0,278	0,189	0,100	0,044	0,011

Pelo histograma,

$$\text{proporção de observações inferior a 9} \approx 0{,}01 + 0{,}01 + 0{,}12 + 0{,}23 = 0{,}37 \quad \left(\text{valor exato} = \frac{34}{90} = 0{,}378\right)$$

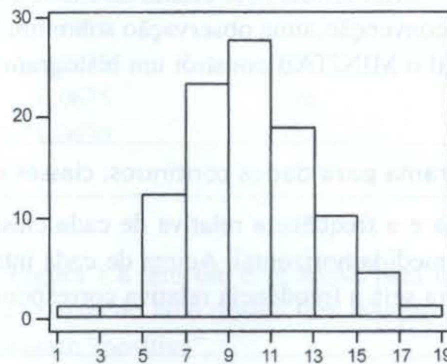


Figura 1.9 Histograma dos dados do consumo de energia do Exemplo 1.10

A freqüência relativa da classe 9–<11 é cerca de 0,27, de forma que estimamos que metade desse valor, ou 0,135, esteja entre 9 e 10. Conseqüentemente,

$$\text{proporção de observações inferior a 10} \approx 0{,}37 + 0{,}135 = 0{,}505 \text{ (pouco mais de 50\%)}$$

O valor exato dessa proporção é 47/90 = 0,522. ■

Não há regras rápidas e absolutas sobre o número de classes ou a escolha das classes em si. Entre 5 e 20 classes serão satisfatórias para a maior parte dos conjuntos de dados. Normalmente, quanto maior o número de observações em um conjunto de dados, mais classes devem ser usadas. Uma regra prática razoável é

$$\text{número de classes} \approx \sqrt{\text{número de observações}}$$

Classes de mesma largura podem não ser uma boa escolha se o conjunto de dados "se estender" para um lado ou para o outro. A Figura 1.10 exibe um gráfico de pontos desse conjunto de dados. Usar um número pequeno de classes de mesma largura resulta em quase todas as observações estarem em apenas uma ou duas classes. Se for usado um grande número de classes de mesma largura, muitas classes terão freqüência zero. Uma opção melhor é usar alguns intervalos mais amplos próximos às observações dos extremos e intervalos mais estreitos na região de alta concentração.

Construindo um histograma para dados contínuos: classes de larguras diferentes

Após determinar as freqüências e as freqüências relativas, calcule a altura de cada retângulo, usando a fórmula

$$\text{altura do retângulo} = \frac{\text{freqüência relativa da classe}}{\text{largura da classe}}$$

As alturas resultantes dos retângulos normalmente são denominadas *densidades* e a escala vertical é a **escala de densidade**. Essa recomendação também funcionará quando as larguras das classes forem iguais.

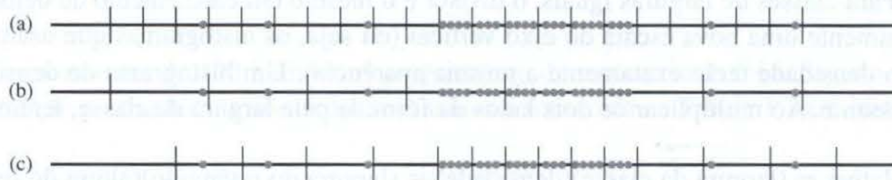


Figura 1.10 Seleção de intervalos de classe para pontos "estendidos": (a) muitos intervalos estreitos de mesma largura; (b) poucos intervalos mais amplos; (c) intervalos de larguras diferentes

Exemplo 1.11

A corrosão das barras de aço da armação é um problema sério em estruturas de concreto localizadas em ambientes afetados por condições climáticas extremas. Por esse motivo, os pesquisadores têm investigado a utilização de barras de reforço feitas de material composto. Um estudo foi executado para desenvolver diretrizes sobre a aderência de barras plásticas reforçadas com fibra de vidro ao concreto ("Design Recommendations for Bond of GFRP Rebars to Concrete," *J. of Structural Engr.*, 1996, p. 247-254). Considere as 48 observações da resistência da aderência medida:

11,5	12,1	9,9	9,3	7,8	6,2	6,6	7,0	13,4	17,1	9,3	5,6
5,7	5,4	5,2	5,1	4,9	10,7	15,2	8,5	4,2	4,0	3,9	3,8
3,6	3,4	20,6	25,5	13,8	12,6	13,1	8,9	8,2	10,7	14,2	7,6
5,2	5,5	5,1	5,0	5,2	4,8	4,1	3,8	3,7	3,6	3,6	3,6

Classe	2–<4	4–<6	6–<8	8–<12	12–<20	20–<30
Freqüência	9	15	5	9	8	2
Freqüência relativa	0,1875	0,3125	0,1042	0,1875	0,1667	0,0417
Densidade	0,094	0,156	0,052	0,047	0,021	0,004

O histograma resultante é exibido na Figura 1.11. A cauda direita ou superior se estende muito além da esquerda ou inferior – um desvio substancial da simetria.

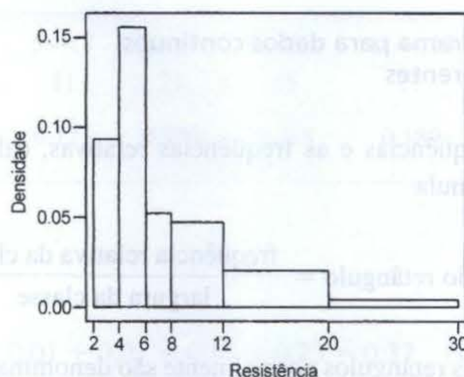


Figura 1.11 Um histograma de densidade em MINITAB dos dados de resistência à aderência do Exemplo 1.11 ■

Quando as larguras de classe são diferentes, deixar de usar uma escala de densidade gera um gráfico com áreas distorcidas. Para classes de larguras iguais, o divisor é o mesmo em cada cálculo de densidade e o cálculo adicional é simplesmente uma nova escala do eixo vertical (ou seja, os histogramas que usam freqüência relativa e os que usam densidade terão exatamente a mesma aparência). Um histograma de densidade possui uma propriedade interessante. Ao multiplicar os dois lados da fórmula pela largura da classe, teremos

$$\text{freqüência relativa} = (\text{largura da classe})(\text{densidade}) = (\text{largura do retângulo})(\text{altura do retângulo})$$
$$= \text{área do retângulo}$$

Isto é, *a área de cada retângulo é a freqüência relativa da classe correspondente.* Além disso, como a soma das freqüências relativas deve ser 1,0 (exceto por arredondamento), *a área total de todos os retângulos em um histograma de densidade é 1.* Sempre é possível desenhar um histograma de forma que a área seja igual à freqüência relativa (isso também é verdade para um histograma de dados discretos). É só usar a escala de densidade. Essa propriedade terá um papel importante na criação de modelos de distribuições no Capítulo 4.

Formatos de histogramas

Os histogramas podem ter diversos formatos. Um histograma **unimodal** é aquele que possui um aclive para um único pico e depois um declive. Um histograma **bimodal** possui dois picos diferentes. A bimodalidade pode ocorrer quando o conjunto de dados consistir em observações sobre dois tipos bastante diferentes de indivíduos ou objetos. Por exemplo: considere um grande conjunto de dados formado por tempos de viagem de automóveis entre San Luis Obispo, Califórnia, e Monterey, Califórnia (excluindo tempo de parada para apreciar a vista, comer etc.). Esse histograma mostraria dois picos: um para os carros que tomaram a rota do interior (cerca de 2,5 horas) e outro para os carros que foram pelo litoral (3,5–4 horas). A bimodalidade, entretanto, não acontece automaticamente nessas situações. A bimodalidade ocorrerá no histograma de dados combinados somente se os dois histogramas separados estiverem "distantes" em relação às suas dispersões. Assim, um grande conjunto de dados consistindo em alturas de alunos de faculdades não deve resultar em um histograma bimodal porque a altura típica dos homens, cerca de 69 polegadas, não está suficientemente distante da altura típica das mulheres, cerca de 64-65 polegadas. Um histograma com mais de dois picos é denominado **multimodal.** Claro que o número de picos pode depender da escolha dos intervalos de classe, particularmente com um pequeno número de observações. Quanto maior o número de classes, maior é a probabilidade de a bimodalidade ou de a multimodalidade se manifestar.

Um histograma é **simétrico** se a metade esquerda for uma imagem refletida da metade direita. Um histograma unimodal tem **inclinação positiva** se a cauda direita ou superior for estendida em comparação à cauda esquerda ou inferior e **inclinação negativa** desviar-se para a esquerda. A Figura 1.12 exibe histogramas "ajustados", obtidos pela sobreposição de uma curva ajustada sobre os retângulos, que ilustram as diversas possibilidades.

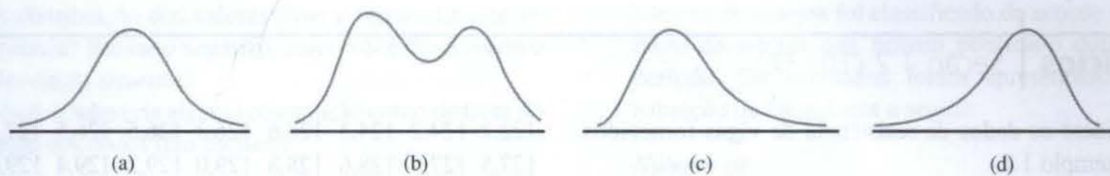


Figura 1.12 Histogramas ajustados: (a) unimodal simétrico; (b) bimodal; (c) desvio positivo; (d) desvio negativo

Dados qualitativos

Tanto uma distribuição de freqüência como um histograma podem ser construídos quando o conjunto de dados for de natureza *qualitativa* (categorizada). Em alguns casos, haverá uma organização natural das classes, por exemplo: calouros, segundanistas, terceiranistas, formandos e graduados, enquanto em outros casos a organização será arbitrária, por exemplo: católicos, judeus, protestantes e assim por diante. Com esses dados categorizados, os intervalos sobre os quais os retângulos serão construídos devem ter a mesma largura.

Exemplo 1.12

Cada membro de uma amostra de 120 indivíduos proprietários de motocicletas foi indagado sobre a marca de sua moto. A distribuição de freqüência dos dados resultantes é fornecida na Tabela 1.2 e o histograma é exibido na Figura 1.13.

Tabela 1.2 Distribuição de freqüência dos dados de motocicletas

Fabricante	Freqüência	Freqüência Relativa
1. Honda	41	0,34
2. Yamaha	27	0,23
3. Kawasaki	20	0,17
4. Harley-Davidson	18	0,15
5. BMW	3	0,03
6. Outro	11	0,09
	120	1,01

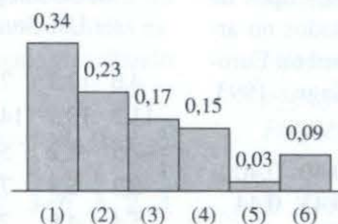


Figura 1.13 Histograma dos dados de motocicletas ■

Dados Multivariados

As técnicas apresentadas até agora referem-se exclusivamente a situações em que cada observação de um conjunto de dados é um único número ou uma única categoria. Os dados, entretanto, freqüentemente são de natureza multivariada. Isto é, se obtivermos uma amostra de indivíduos ou objetos e em cada um tivermos duas ou mais medidas, cada "observação" consistirá em diversas medidas de um indivíduo ou objeto. A amostra é bivariada se cada observação consistir em duas medidas ou respostas, de forma que o conjunto de dados possa ser representado como $(x_1, y_1), ..., (x_n, y_n)$. Por exemplo: x pode se referir ao tamanho do motor e y ao seu deslocamento; ou x pode se referir à marca da calculadora de um formando e y à sua formação acadêmica. Nos capítulos 11-14, analisaremos conjuntos de dados multivariados desse tipo; assim, adiaremos uma discussão mais detalhada até lá.

Exercícios | Seção 1.2 (10–32)

10. Considere os dados de resistência de vigas fornecidos no Exemplo 1.2.

a. Construa um diagrama de caule e folha dos dados. O que parece ser um valor de resistência representativo? As observações parecem estar concentradas ao redor do valor representativo ou dispersas?

b. O diagrama parece razoavelmente simétrico ao redor de um valor representativo ou seu formato pode ser descrito de outra forma?

c. Parece haver valores aberrantes de resistência?

d. Que proporção de observações de resistência dessa amostra excedem 10 MPa?

11. Cada pontuação do conjunto de notas de um exame a seguir está nas dezenas 60, 70, 80 ou 90. Um diagrama de caule e folha com apenas os quatro caules 6, 7, 8 e 9 não forneceria uma descrição muito detalhada da distribuição das pontuações. Nessas situações, é desejável usarmos caules repetidos. Aqui podemos repetir o caule 6 duas vezes, usando 6L para pontuações na parte inferior da dezena dos 60 (folhas 0, 1, 2, 3 e 4) e 6H para as pontuações na parte superior da dezena dos 60 (folhas 5, 6, 7, 8 e 9). De forma similar, os outros caules podem ser repetidos duas vezes para obtermos um diagrama consistindo em oito linhas. Construa esse diagrama para as pontuações fornecidas. Que característica dos dados é realçada por ele?

74	89	80	93	64	67	72	70	66	85	89	81	81
71	74	82	85	63	72	81	81	95	84	81	80	70
69	66	60	83	85	98	84	68	90	82	69	72	87
88												

12. Os dados anexos de densidade para diversos tipos de madeira usados em construção foram relatados no artigo "Bolted Connection Design Values Based on European Yield Model" (*J. of Structural Engr.*, 1993, p. 2169-2186):

0,31	0,35	0,36	0,36	0,37	0,38	0,40	0,40	0,40
0,41	0,41	0,42	0,42	0,42	0,42	0,42	0,43	0,44
0,45	0,46	0,46	0,47	0,48	0,48	0,48	0,51	0,54
0,54	0,55	0,58	0,62	0,66	0,66	0,67	0,68	0,75

Construa um diagrama de caule e folha usando caules repetidos (veja o exercício anterior) e comente suas características interessantes.

13. As propriedades mecânicas permissíveis para projetos estruturais de veículos aeroespaciais metálicos exigem um método aprovado para análise estatística de dados de testes empíricos. O artigo "Establishing Mechanical Property Allowables for Metals" (*J. of Testing and Evaluation*, 1998, p. 293-299) usou os dados sobre resistência à tração (ksi) como base para definir as dificuldades de desenvolvimento do método.

122,2	124,2	124,3	125,6	126,3	126,5	126,5	127,2	127,3
127,5	127,9	128,6	128,8	129,0	129,2	129,4	129,6	130,2
130,4	130,8	131,3	131,4	131,4	131,5	131,6	131,6	131,8
131,8	132,3	132,4	132,4	132,5	132,5	132,5	132,5	132,6
132,7	132,9	133,0	133,1	133,1	133,1	133,1	133,2	133,2
133,2	133,3	133,3	133,5	133,5	133,5	133,8	133,9	134,0
134,0	134,0	134,0	134,1	134,2	134,3	134,4	134,4	134,6
134,7	134,7	134,7	134,8	134,8	134,8	134,9	134,9	135,2
135,2	135,2	135,3	135,3	135,4	135,5	135,5	135,6	135,6
135,7	135,8	135,8	135,8	135,8	135,8	135,9	135,9	135,9
135,9	136,0	136,0	136,1	136,2	136,2	136,3	136,4	136,4
136,6	136,8	136,9	136,9	137,0	137,1	137,2	137,6	137,6
137,8	137,8	137,8	137,9	137,9	138,2	138,2	138,3	138,3
138,4	138,4	138,4	138,5	138,5	138,6	138,7	138,7	139,0
139,1	139,5	139,6	139,8	139,8	140,0	140,0	140,7	140,7
140,9	140,9	141,2	141,4	141,5	141,6	142,9	143,4	143,5
143,6	143,8	143,8	143,9	144,1	144,5	144,5	147,7	147,7

a. Construa um diagrama de caule e folha dos dados excluindo (truncando) inicialmente os dígitos decimais e depois repetindo cada caule cinco vezes (uma vez para as folhas 1 e 2, uma segunda vez para as folhas 3 e 4 etc.). Por que é relativamente fácil identificar um valor de resistência representativo?

b. Construa um histograma usando classes de mesma largura em que a primeira classe possua um limite inferior a 122 e um limite superior a 124. Comente, então, características interessantes do histograma.

14. O conjunto de dados a seguir consiste de observações da vazão de chuveiros (L/min) de uma amostra de $n = 129$ lares em Perth, Austrália ("An Application of Bayes Methodology to the Analysis of Diary Records in a Water Use Study", *J. Amer. Stat. Assoc.*, 1987, p. 705-711):

4,6	12,3	7,1	7,0	4,0	9,2	6,7	6,9	11,5	5,1
11,2	10,5	14,3	8,0	8,8	6,4	5,1	5,6	9,6	7,5
7,5	6,2	5,8	2,3	3,4	10,4	9,8	6,6	3,7	6,4
8,3	6,5	7,6	9,3	9,2	7,3	5,0	6,3	13,8	6,2
5,4	4,8	7,5	6,0	6,9	10,8	7,5	6,6	5,0	3,3
7,6	3,9	11,9	2,2	15,0	7,2	6,1	15,3	18,9	7,2
5,4	5,5	4,3	9,0	12,7	11,3	7,4	5,0	3,5	8,2
8,4	7,3	10,3	11,9	6,0	5,6	9,5	9,3	10,4	9,7
5,1	6,7	10,2	6,2	8,4	7,0	4,8	5,6	10,5	14,6
10,8	15,5	7,5	6,4	3,4	5,5	6,6	5,9	15,0	9,6
7,8	7,0	6,9	4,1	3,6	11,9	3,7	5,7	6,8	11,3
9,3	9,6	10,4	9,3	6,9	9,8	9,1	10,6	4,5	6,2
8,3	3,2	4,9	5,0	6,0	8,2	6,3	3,8	6,0	

a. Construa um diagrama de caule e folha dos dados.

b. Qual taxa é considerada vazão típica ou representativa?

c. O diagrama parece ser concentrado ou disperso?

d. A distribuição dos valores parece razoavelmente simétrica? Em caso negativo, como você descreveria o desvio da simetria?

e. Você descreveria alguma observação como distante do resto dos dados (um *outlier*)?

15. Um artigo da *Consumer Reports* sobre pasta de amendoim (setembro de 1990) relatou as seguintes pontuações para diversas marcas:

Cremosa	56	44	62	36	39	53	50	65	45	40
	56	68	41	30	40	50	56	30	22	
Crocante	62	53	75	42	47	40	34	62	52	
	50	34	42	36	75	80	47	56	62	

Construa um diagrama de caule e folha comparativo, relacionando caules na parte central da página e, então, exibindo as folhas de cremosa à direita e as de crocante à esquerda. Descreva semelhanças e diferenças para os dois tipos.

16. O artigo citado no Exemplo 1.2 também fornece as informações de resistência de corpos de prova a seguir:

6,1	5,8	7,8	7,1	7,2	9,2	6,6	8,3	7,0	8,3
7,8	8,1	7,4	8,5	8,9	9,8	9,7	14,1	12,6	11,2

a. Construa um diagrama de caule e folha comparativo (veja o exercício anterior) dos dados de vigas e cilindros e, então, responda às questões nas partes (b)-(d) do Exercício 10 para as observações sobre os corpos de prova.

b. De que formas os dois lados do diagrama são similares? Há diferenças óbvias entre as observações sobre vigas e corpos de prova?

c. Construa um gráfico de pontos com os dados dos corpos de prova.

17. Os transdutores de temperatura de um determinado tipo são enviados em lotes de 50. Uma amostra de 60 lotes foi selecionada e o número de transdutores fora das especificações em cada lote foi determinado, resultando nos dados a seguir:

2	1	2	4	0	1	3	2	0	5	3	3	1	3	2	4	7	0	2	3
0	4	2	1	3	1	1	3	4	1	2	3	2	2	8	4	5	1	3	1
5	0	2	3	2	1	0	6	4	2	1	6	0	3	3	3	6	1	2	3

a. Determine as freqüências e freqüências relativas dos valores observados de x = número de transdutores fora das especificações em um lote.

b. Que proporção de lotes na amostra possui no máximo cinco transdutores fora das especificações? Que proporção tem menos de cinco? Que proporção possui no mínimo cinco unidades fora das especificações?

c. Desenhe um histograma dos dados, usando a freqüência relativa na escala vertical e comente suas características.

18. Em um estudo de produtividade literária ("Lotka's Test," *Collection Mgmt.*, 1982, p. 111-118), um grande número de autores foi classificado de acordo com o número de artigos que tinham publicado durante certo período. Os resultados foram apresentados na distribuição de freqüência a seguir:

Número de artigos	1	2	3	4	5	6	7	8
Freqüência	784	204	127	50	33	28	19	19

Número de artigos	9	10	11	12	13	14	15	16	17
Freqüência	6	7	6	7	4	4	5	3	3

a. Construa um histograma correspondente a essa distribuição de freqüência. Qual é a característica mais interessante do formato da distribuição?

b. Que proporção desses autores publicou no mínimo cinco artigos? No mínimo 10 artigos? Mais de 10 artigos?

c. Suponha que os cinco que publicaram 15 artigos, os três que publicaram 16 e os três que publicaram 17 tenham sido agrupados em uma única categoria exibida como "≥15." Você pode construir um histograma? Explique.

d. Suponha que, em vez dos valores 15, 16 e 17 relacionados separadamente, eles tenham sido combinados em uma categoria 15-17 com freqüência 11. Você pode construir um histograma? Explique.

19. O número de partículas de contaminação de uma pastilha de silício antes de certo processo de limpeza foi determinado para cada pastilha em uma amostra de tamanho 100, resultando nas freqüências a seguir:

Número de partículas	0	1	2	3	4	5	6	7
Freqüência	1	2	3	12	11	15	18	10

Número de partículas	8	9	10	11	12	13	14
Freqüência	12	4	5	3	1	2	1

a. Que proporção das pastilhas da amostra tinha ao menos uma partícula? Ao menos cinco partículas?

b. Que proporção das pastilhas da amostra tinha entre cinco e 10 (inclusive) partículas? Estritamente entre cinco e 10 partículas?

c. Desenhe um histograma usando a freqüência relativa no eixo vertical. Como você descreveria o formato do histograma?

20. O artigo "Determination of Most Representative Subdivision" (*J. of Energy Engr.*, 1993, p. 43-55) forneceu dados sobre diversas características de subdivisões que podiam ser usadas na decisão de fornecimento de energia via linhas subterrâneas ou aéreas. Seguem os valores da variável x = comprimento total das ruas dentro de uma subdivisão:

1280	5320	4390	2100	1240	3060	4770
1050	360	3330	3380	340	1000	960
1320	530	3350	540	3870	1250	2400
960	1120	2120	450	2250	2320	2400
3150	5700	5220	500	1850	2460	5850
2700	2730	1670	100	5770	3150	1890
510	240	396	1419	2109		

a. Construa um diagrama de caule e folha, usando o dígito de milhares como caule e o de centenas como folha, e comente suas características.

b. Construa um histograma, usando limites de classe de 0, 1000, 2000, 3000, 4000, 5000 e 6000. Que proporção de subdivisões possui comprimento total inferior a 2000? Entre 2000 e 4000? Como você descreveria o formato do histograma?

21. O artigo citado no Exercício 20 também forneceu os seguintes valores das variáveis y = números de *culs-de-sac* e z = número de interseções:

y	1	0	1	0	0	2	0	1	1	1	2	1	0	0	1	1	0	1	1
z	1	8	6	1	1	5	3	0	0	4	4	0	0	1	2	1	4	0	4
y	1	1	0	0	0	1	1	2	0	1	2	2	1	1	0	2	1	1	0
z	0	3	0	1	1	0	1	3	2	4	6	6	0	1	1	8	3	3	5
y	1	5	0	3	0	1	1	0	0										
z	0	5	2	3	1	0	0	0	3										

a. Construa um histograma dos dados de y. Que proporção dessas subdivisões não possuíam *culs-de-sac*? Ao menos um *cul-de-sac*?

b. Construa um histograma para os dados z. Que proporção dessas subdivisões têm no máximo cinco interseções? Menos de cinco interseções?

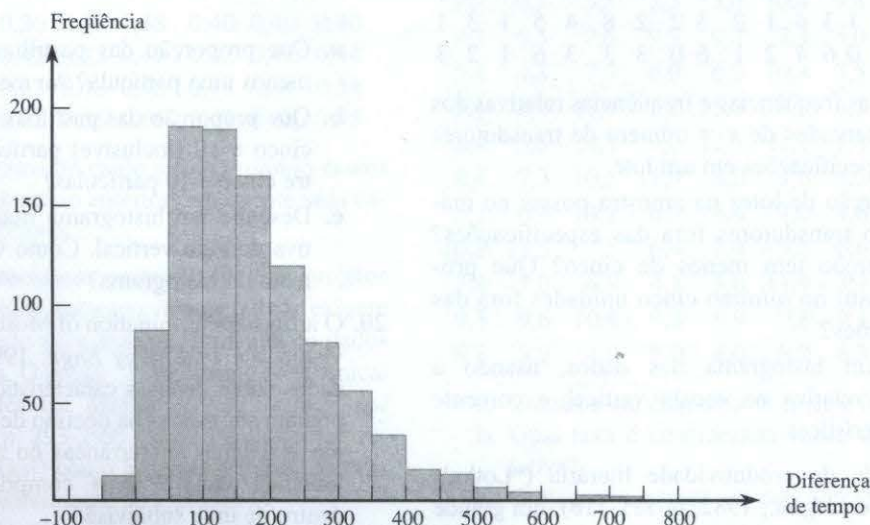
22. Como varia a velocidade de um corredor no curso de uma maratona (uma distância de 42,195 km)? Considere a determinação do tempo de corrida dos cinco primeiros quilômetros e o tempo de corrida entre os pontos dos kms 35 e 40; subtraia o primeiro tempo do último. Um valor positivo dessa diferença corresponde a um corredor que diminui seu ritmo no final da corrida. O histograma a seguir se baseia nos tempos de corredores que participaram de diversas maratonas diferentes no Japão ("Factors Affecting Runners' Marathon Performance", *Chance*, Fall, 1993, p. 24-30). Que características são interessantes nesse histograma? Qual é um valor típico da diferença ? *Grosso modo*, que proporção dos maratonistas correu a última etapa mais rápido que a primeira?

23. Em um estudo de quebras de urdidura durante a tecelagem de tecidos (*Technometrics*, 1982, p. 63), 100 amostras de fios foram testadas. O número de ciclos de esforço para quebra foi determinado para cada amostra de fio, resultando nos dados a seguir:

86	146	251	653	98	249	400	292	131	169
175	176	76	264	15	364	195	262	88	264
157	220	42	321	180	198	38	20	61	121
282	224	149	180	325	250	196	90	229	166
38	337	65	151	341	40	40	135	597	246
211	180	93	315	353	571	124	279	81	186
497	182	423	185	229	400	338	290	398	71
246	185	188	568	55	55	61	244	20	284
393	396	203	829	239	236	286	194	277	143
198	264	105	203	124	137	135	350	193	188

a. Construa um histograma de freqüência relativa com base nos intervalos de classe 0–<100, 100–<200, ... e comente as características do histograma.

Histograma do Exercício 22



b. Construa um histograma com base nos seguintes intervalos de classe: 0–<50, 50–<100, 100–<150, 150–<200, 200–<300, 300–<400, 400–<500, 500–<600 e 600–<900.

c. Se as especificações de tecelagem exigem um esforço de quebra de ao menos 100 ciclos, que proporção das amostras de fio dessa amostra deve ser considerada satisfatória?

24. O conjunto de dados anexo consiste em observações da resistência de corte (lb) de soldas de ponto ultrassônicas feitas sobre um determinado tipo de chapa de Alclad. Construa um histograma de freqüência relativa com base em 10 classes de mesma largura com limites 4000, 4200, ... [O histograma coincide com o mostrado em "Comparison of Properties of Joints Prepared by Ultrasonic Welding and Other Means" (*J. of Aircraft*, 1983, p. 552-556).] Comente as características.

5434	4948	4521	4570	4990	5702	5241
5112	5015	4659	4806	4637	5670	4381
4820	5043	4886	4599	5288	5299	4848
5378	5260	5055	5828	5218	4859	4780
5027	5008	4609	4772	5133	5095	4618
4848	5089	5518	5333	5164	5342	5069
4755	4925	5001	4803	4951	5679	5256
5207	5621	4918	5138	4786	4500	5461
5049	4974	4592	4173	5296	4965	5170
4740	5173	4568	5653	5078	4900	4968
5248	5245	4723	5275	5419	5205	4452
5227	5555	5388	5498	4681	5076	4774
4931	4493	5309	5582	4308	4823	4417
5364	5640	5069	5188	5764	5273	5042
5189	4986					

25. A transformação de valores de dados por meio de uma função matemática, como $\sqrt{x}$ ou $1/x$, normalmente resulta em um conjunto de números com "melhores" propriedades estatísticas do que os dados originais. Em particular, é possível encontrar uma função para a qual o histograma dos valores transformados seja mais simétrico (ou, melhor ainda, mais próximo de uma curva normal) do que os dados originais. Como exemplo, o artigo "Time Lapse Cinematographic Analysis of BerylliumLung Fibroblast Interactions" (*Environ. Research*, 1983, p. 34-43) relatou os resultados de experimentos projetados para estudar o comportamento de algumas células que foram expostas ao berílio. Uma característica importante de tal célula individual é seu tempo de interdivisão (IDT). Os IDTs foram determinados para um grande número de células em condições de exposição (tratamento) e não-exposição (controle). Os autores do artigo usaram uma transformação logarítmica, isto é, valor transformado = log (valor original). Considere os seguintes dados representativos de IDT:

IDT	$\log_{10}$(IDT)	IDT	$\log_{10}$(IDT)	IDT	$\log_{10}$(IDT)
28,1	1,45	60,1	1,78	21,0	1,32
31,2	1,49	23,7	1,37	22,3	1,35
13,7	1,14	18,6	1,27	15,5	1,19
46,0	1,66	21,4	1,33	36,3	1,56
25,8	1,41	26,6	1,42	19,1	1,28
16,8	1,23	26,2	1,42	38,4	1,58
34,8	1,54	32,0	1,51	72,8	1,86
62,3	1,79	43,5	1,64	48,9	1,69
28,0	1,45	17,4	1,24	21,4	1,33
17,9	1,25	38,8	1,59	20,7	1,32
19,5	1,29	30,6	1,49	57,3	1,76
21,1	1,32	55,6	1,75	40,9	1,61
31,9	1,50	25,5	1,41		
28,9	1,46	52,1	1,72		

Use os intervalos de classes 10–<20, 20–<30.... para construir um histograma dos dados originais. Use os intervalos 1,1–<1,2, 1,2–<1,3, ... para fazer o mesmo para os dados transformados. Qual é o efeito da transformação?

26. O índice de céu claro foi determinado para o céu de Bagdá, compreendendo cada um dos 365 dias de um dado ano ("Contribution to the Study of the Solar Radiation Climate of the Baghdad Environment", *Solar Energy*, 1990, p. 7-12). A tabela a seguir fornece os resultados.

Classe	Freqüência
0,15–<0,25	8
0,25–<0,35	14
0,35–<0,45	28
0,45–<0,50	24
0,50–<0,55	39
0,55–<0,60	51
0,60–<0,65	106
0,65–<0,70	84
0,70–<0,75	11

a. Determine as freqüências relativas e desenhe o histograma correspondente.

b. Dias nublados são aqueles com o índice de céu limpo inferior a 0,35. Em que porcentagem dos dias o céu esteve nublado?

c. Dias de céu claro são aqueles para os quais o índice é no mínimo 0,65. Em que porcentagem dos dias o céu esteve limpo?

27. O artigo "Study on the Life Distribution of Microdrills" (*J. of Engr. Manufacture*, 2002: 301-305) relatou as observações a seguir, relacionadas em ordem crescente, da vida útil das brocas (número de furos que uma broca faz antes de quebrar), quando os furos são feitos em uma determinada liga de bronze.

11	14	20	23	31	36	39	44	47	50
59	61	65	67	68	71	74	76	78	79
81	84	85	89	91	93	96	99	101	104
105	105	112	118	123	136	139	141	148	158
161	168	184	206	248	263	289	322	388	513

a. Por que uma distribuição de freqüência não pode ter por base os intervalos de classe 0–50, 50–100, 100–150 e assim por diante?

b. Construa uma distribuição de freqüência e um histograma dos dados usando limites de classes 0, 50, 100, ... e então faça comentários sobre as características interessantes.

c. Construa uma distribuição de freqüência e um histograma dos logaritmos naturais relacionados às observações de vida útil e comente as características interessantes.

d. Que proporção das observações de vida útil dessa amostra é inferior a 100? Que proporção das observações é igual ou maior que 200?

28. Construa um gráfico de pontos para a série de dados anexa. Os dados são mensais e foram obtidos durante o período de 1985-1989. Cada valor é a radiação solar média na faixa 385-530 nm como porcentagem da radiação total ("Global Energy in the Different Spectral Bands at Dhahran, Saudi Arabia," *J. Solar Energy Engr.*, 1991, p. 290-294). Comente sobre algumas características interessantes dos dados.

20,9	19,6	20,4	20,3	20,8	20,6	20,5	20,4
19,9	19,8	19,5	20,2	16,5	18,3	18,7	19,6
20,0	20,0	19,5	19,6	19,1	18,8	18,3	17,6
17,2	17,8	18,7	19,0	19,0	18,6	18,8	19,0
18,5	18,3	17,5	16,9	17,0	17,8	18,1	18,8
18,9	18,9	19,1	18,8	18,4	17,8	17,0	16,8
17,9	18,4	19,0	19,4	19,7	19,5	19,5	19,5
19,0	18,7	18,1	17,9				

29. Considere os dados a seguir sobre os tipos de queixas de saúde (J = inflamação de articulações, F = fadiga, B = dor nas costas, M = fadiga muscular, C = tosse, N = irritação nasal/coriza, O = outros) feitas por agricultores. Obtenha as freqüências e as freqüências relativas das diversas categorias e desenhe um histograma. (Os dados são consistentes com as porcentagens fornecidas no artigo "Physiological Effects of Work Stress and Pesticide Exposure in Tree Planting by British Columbia Silviculture Workers," *Ergonomics*, 1993, p. 951-961.)

O	O	N	J	C	F	B	B	F	O	J	O	O	M
O	F	F	O	O	N	O	N	J	F	J	B	O	C
J	O	J	J	F	N	O	B	M	O	J	M	O	B
O	F	J	O	O	B	N	C	O	O	O	M	B	F
J	O	F	N										

30. Um **Diagrama de Pareto** é uma variação de um histograma para dados categorizados resultantes de um estudo de controle de qualidade. Cada categoria representa um tipo diferente de não-conformidade de produto ou problema de produção. As categorias são ordenadas de forma que aquela com maior freqüência seja exibida na extremidade esquerda, seguida pela categoria com a segunda maior freqüência e assim por diante. Suponha que as informações a seguir tenham sido obtidas sobre não-conformidades em pacotes de circuitos: componentes com falha, 126; componentes incorretos, 210; soldas insuficientes, 67; soldas em excesso, 54; falta de componentes, 131. Construa um Diagrama de Pareto.

31. A **freqüência acumulada** e a freqüência relativa acumulada de um determinado intervalo de classe são a soma das freqüências e freqüências relativas, respectivamente, desse intervalo e de todos os intervalos abaixo dele. Se, por exemplo, houver quatro intervalos com freqüências 9, 16, 13 e 12, as freqüências acumuladas serão 9, 25, 38 e 50 e as freqüências relativas acumuladas serão 0,18, 0,50, 0,76 e 1,00. Calcule as freqüências acumuladas e as freqüências relativas acumuladas para os dados do Exercício 24.

32. Uma carga de incêndio (MJ/m^2) é a energia térmica que pode ser liberada por metro quadrado de área de piso pela combustão de seu conteúdo e da estrutura em si. O artigo "Fire Loads in Office Buildings" (*J. of Structural Engr.*, 1997, p. 365-368) forneceu as seguintes porcentagens acumuladas (lidas de um gráfico) relativas a cargas de incêndio em uma amostra de 388 salas:

Valor	0	150	300	450	600
% Acumulada	0	19,3	37,6	62,7	77,5
Valor	750	900	1050	1200	1350
% Acumulada	87,2	93,8	95,7	98,6	99,1
Valor	1500	1650	1800	1950	
% Acumulada	99,5	99,6	99,8	100,0	

a. Construa um histograma de freqüência relativa e comente as características interessantes.

b. Que proporção das cargas de incêndio é inferior a 600? Maior ou igual a 1200?

c. Que proporção das cargas está entre 600 e 1200?

1.3 Medidas de localização

Os resumos visuais de dados são excelentes ferramentas para obter impressões e idéias iniciais. Uma análise mais formal de dados freqüentemente exige o cálculo e a interpretação de medidas-resumo numéricas simples. Isto é, a partir dos dados, tentamos extrair diversos números simples, que servem para caracterizar o conjunto de

dados e indicar algumas informações consideráveis. Nossa preocupação principal será com os dados numéricos. Alguns comentários sobre dados categorizados serão apresentados no final da seção.

Suponha, então, que nosso conjunto de dados é do formato $x_1, x_2, ..., x_n$, onde cada x_i é um número. Que características de tal conjunto de números são de maior interesse e merecem ênfase? Uma característica importante de um conjunto de números é sua localização e, em particular, seu centro. Esta seção apresenta métodos de descrição da localização de um conjunto de dados. Na Seção 1.4, apresentaremos os métodos de medida da dispersão de um conjunto de números.

A média

Para um determinado conjunto de números $x_1, x_2, ..., x_n$, a medida mais familiar e útil do centro é a *média* do conjunto. Como quase sempre temos os vários x_i constituindo uma amostra, freqüentemente chamaremos a média aritmética de *média amostral* e a representaremos por $\bar{x}$.

DEFINIÇÃO A **média amostral** $\bar{x}$ das observações $x_1, x_2, ..., x_n$, é dada por

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

O numerador de $\bar{x}$ pode ser escrito mais informalmente como $\sum x_i$, onde a soma se dá sobre todas as observações da amostra.

Para informar $\bar{x}$, recomendamos o uso de precisão decimal de um dígito a mais do que a precisão dos x_i. Dessa forma, se as observações forem distâncias de parada com $x_1 = 125$, $x_2 = 131$ e assim por diante, podemos ter $\bar{x} = 127,3$ pés.

Exemplo 1.13

As trincas em aço e ferro causadas por fadiga de corrosão cáustica foram estudadas em decorrência de falhas em rebites de caldeiras de aço e em rotores a vapor. Considere as observações a seguir sobre x = comprimento da trinca (μm) como resultado de testes de fadiga por corrosão devido a cargas constantes em amostras de barras de tração lisas durante um período de tempo fixo. (Os dados são consistentes com um histograma e as quantidades-resumo do artigo "On the Role of Phosphorus in the Caustic Stress Corrosion Cracking of Low Alloy Steels", *Corrosion Science*, 1989: 53-68.)

$x_1 = 16,1$ $x_2 = 9,6$ $x_3 = 24,9$ $x_4 = 20,4$ $x_5 = 12,7$ $x_6 = 21,2$ $x_7 = 30,2$
$x_8 = 25,8$ $x_9 = 18,5$ $x_{10} = 10,3$ $x_{11} = 25,3$ $x_{12} = 14,0$ $x_{13} = 27,1$ $x_{14} = 45,0$
$x_{15} = 23,3$ $x_{16} = 24,2$ $x_{17} = 14,6$ $x_{18} = 8,9$ $x_{19} = 32,4$ $x_{20} = 11,8$ $x_{21} = 28,5$

A Figura 1.14 mostra um diagrama de caule e folha dos dados. Um comprimento de trinca no início da faixa dos 20 parece ser "típica".

0H	96 89
1L	27 03 40 46 18
1H	61 85
2L	49 04 12 33 42
2H	58 53 71 85
3L	02 24
3H	
4L	
4H	50

Caule: dígito das dezenas
Folha: dígito das unidades e das dezenas

Figura 1.14 Um diagrama de caule e folha dos dados dos comprimentos de trincas

Sendo $\sum x_i = 444,8$, a média amostral é

$$\bar{x} = \frac{444,8}{21} = 21,18$$

um valor consistente com as informações ilustradas pelo diagrama de caule e folha. ■

Uma interpretação física de $\bar{x}$ demonstra como ela mede a localização (centro) de uma amostra. Imagine desenhar e definir a escala em um eixo horizontal e depois represente cada observação da amostra por um peso de uma libra colocado no ponto correspondente no eixo. O único ponto em que pode ser colocado um apoio para equilibrar o sistema de pesos é o correspondente ao valor de $\bar{x}$ (veja a Figura 1.15).

Da mesma forma que $\bar{x}$ representa o valor médio das observações de uma amostra, a média de todos os valores da população pode ser calculada. Essa média é denominada **média da população** e é representada pela letra grega μ. Quando houver N valores na população (uma população finita), μ = (somatória dos N valores da população)/N. Nos capítulos 3 e 4, forneceremos uma definição mais geral de μ que se aplica a populações finitas e (conceitualmente) infinitas. Da mesma forma que $\bar{x}$ é uma medida de localização de amostra importante e interessante, μ é uma característica interessante e importante (freqüentemente a mais importante) de uma população. Nos capítulos sobre inferência estatística, apresentaremos métodos com base na média amostral para obtenção de conclusões sobre a média de uma população. Por exemplo: podemos usar a média amostral $\bar{x} = 21,18$ calculada no Exemplo 1.13 como uma *estimativa de ponto* (um único número que é o "melhor" palpite) de μ, o comprimento médio verdadeiro de todas as amostras tratadas como descrito.

A média sofre de uma deficiência que a torna uma medida de centro inadequada sob algumas circunstâncias: seu valor pode ser bastante afetado pela presença de um único *outlier* (uma observação incomumente grande ou pequena). No Exemplo 1.13, o valor $x_{14} = 45,0$ obviamente é um *outlier*. Sem esta observação, $\bar{x} = 399,8/20 = 19,99$, o *outlier* aumenta a média em mais de 1 μm. Se a observação 45,0 μm fosse substituída pelo valor catastrófico de 295,0 μm, um *outlier* realmente extremo, então $\bar{x} = 694,8/21 = 33,09$, que é maior que todas as observações, exceto uma.

Uma amostra de salários normalmente produz alguns poucos valores aberrantes (dos sortudos que possuem um salário astronômico) e o uso do salário médio como medida de localização freqüentemente será ilusório. Esses exemplos sugerem que procuremos uma medida menos sensível a valores fora da faixa que $\bar{x}$, assim, proporemos uma momentaneamente. Entretanto, apesar de $\bar{x}$ ter essa falha potencial, ela ainda é a medida mais usada, em grande parte porque há muitas populações para as quais um *outlier* extremo na amostra seria altamente improvável. Ao obter uma amostra de uma tal população (a população normal ou em forma de sino, é o exemplo mais importante), a média amostral tenderá a ser estável e muito representativa.

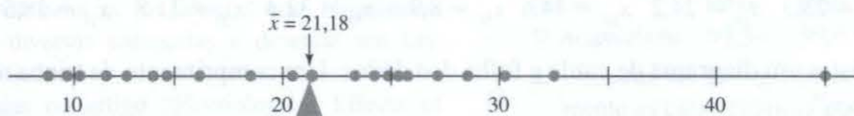


Figura 1.15 A média como ponto de equilíbrio de um sistema de pesos

A mediana

A palavra *mediana* é sinônimo de "metade" e a mediana amostral é o valor do meio quando as observações são ordenadas da menor para a maior. Quando as observações estiverem representadas por $x_1, \ldots, x_n$, usaremos o símbolo $\tilde{x}$ para representar a mediana amostral.

DEFINIÇÃO A **mediana amostral** é obtida pela ordenação das n observações da menor para a maior (com os valores repetidos incluídos, de forma que cada observação da amostra seja exibida na lista ordenada). Assim,

$$\tilde{x} = \begin{cases} \text{O único valor médio se } n \text{ for ímpar} & = \left(\dfrac{n+1}{2}\right) \text{enésimo valor ordenado} \\ \text{A média dos dois valores médios se } n \text{ for par} & = \text{média dos valores ordenados } \left(\dfrac{n}{2}\right) \text{e} \left(\dfrac{n}{2}+1\right) \end{cases}$$

Exemplo 1.14

O risco de desenvolvimento de deficiência de ferro é especialmente alto durante a gravidez. O problema na detecção dessa deficiência é que alguns métodos de determinação de nível de ferro podem ser afetados pelo próprio estado de gravidez. Considere os dados a seguir sobre a concentração do receptor de transferrina de uma amostra de mulheres com evidências laboratoriais de uma visível anemia por deficiência de ferro ("Serum Transferrin Receptor for the Detection of Iron Deficiency in Pregnancy," *Amer. J. of Clinical Nutrition*, 1991: p. 1077-1081):

$$x_1 = 15{,}2 \quad x_2 = 9{,}3 \quad x_3 = 7{,}6 \quad x_4 = 11{,}9 \quad x_5 = 10{,}4 \quad x_6 = 9{,}7$$
$$x_7 = 20{,}4 \quad x_8 = 9{,}4 \quad x_9 = 11{,}5 \quad x_{10} = 16{,}2 \quad x_{11} = 9{,}4 \quad x_{12} = 8{,}3$$

A lista dos valores ordenados é

7,6 8,3 9,3 9,4 9,4 9,7 10,4 11,5 11,9 15,2 16,2 20,4

Como $n = 12$ é par, tiramos a média $n/2$ = *do sexto e sétimo valores ordenados*:

$$\text{mediana amostral} = \frac{9{,}7 + 10{,}4}{2} = 10{,}05$$

Observe que, se a maior observação, 20,4, não tivesse aparecido na amostra, a mediana amostral resultante para as $n = 11$ observações teria sido o único valor médio, 9,7 (o $(n + 1)/2$ = sexto valor ordenado). A média amostral é $\bar{x} = \Sigma x_i/n = 139{,}3/12 = 11{,}61$, que é um pouco maior que a mediana, por causa dos *outliers*, 15,2, 16,2 e 20,4. ■

Os dados do Exemplo 1.14 ilustram uma propriedade importante de $\tilde{x}$ em comparação com $\bar{x}$: a mediana amostral é muito insensível a muitos valores extremamente pequenos ou extremamente grandes. Se, por exemplo, aumentássemos os dois maiores x_i de 16,2 e 20,4 para 26,2 e 30,4, respectivamente, $\tilde{x}$ não seria afetado. Dessa forma, no tratamento de valores de dados fora da faixa, $\bar{x}$ e $\tilde{x}$ são extremidades opostas de um espectro: $\bar{x}$ é sensível mesmo a um único valor, enquanto $\tilde{x}$ é insensível a um grande número de valores fora da faixa.

Como os valores grandes na amostra do Exemplo 1.14 afetam $\bar{x}$ mais que $\tilde{x}$, $\tilde{x} < \bar{x}$ para esses dados. Apesar de $\bar{x}$ e $\tilde{x}$ fornecerem uma medida para o centro da amostra em um conjunto de dados, eles em geral não serão iguais, porque enfocam diferentes aspectos da amostra.

De forma análoga, $\tilde{x}$ como valor médio na amostra é o valor médio da população, a **mediana da população**, representada por $\tilde{\mu}$. Como acontece com $\bar{x}$ e μ, podemos considerar o uso da mediana amostral $\tilde{x}$ para fazer inferências de $\tilde{\mu}$. No Exemplo 1.14, podemos usar $\tilde{x} = 10{,}05$ como estimativa da concentração da mediana em toda a população a partir da qual a amostra foi selecionada. Uma mediana normalmente é usada para descrever dados de salários ou rendimentos (porque ela não é influenciada por alguns grandes salários). Se a

mediana de uma amostra dos salários de engenheiros fosse $\tilde{x}$ = \$ 66.416, poderíamos usá-la como base para concluir que o salário mediano dos engenheiros excede \$ 60.000.

A média da população μ e a mediana $\tilde{\mu}$ normalmente não serão idênticas. Se a distribuição da população tiver desvio positivo ou negativo, conforme ilustrado na Figura 1.16, então $\mu \neq \tilde{\mu}$. Quando esse for o caso, ao fazer inferências, devemos primeiro decidir quais características das populações são de maior interesse e então proceder de acordo.

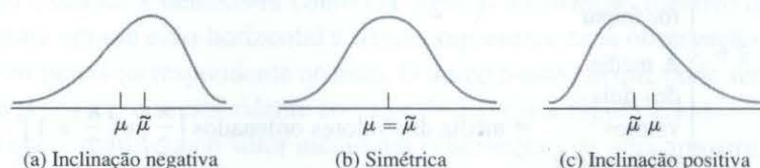


Figura 1.16 Três formatos diferentes para uma distribuição de população

Outras medidas de localização
Quartis, Percentis e Médias Aparadas

A mediana (de população ou amostra) divide o conjunto de dados em duas partes de mesmo tamanho. Para obter melhores medidas de localização, podemos dividir os dados em mais de duas partes. Grosso modo, os quartis dividem o conjunto em quatro partes iguais, sendo que as observações acima do terceiro quartil constituem o quarto superior do conjunto de dados, o segundo quartil é idêntico à mediana e o primeiro quartil separa o quarto inferior dos três quartos superiores. De forma similar, um conjunto de dados (amostra ou população) pode ser dividido mais detalhadamente usando percentis; o 99º percentil separa o 1% superior do restante, e assim por diante. A menos que o número de observações seja um múltiplo de 100, recomenda-se cuidado na utilização de percentis. Usaremos percentis no Capítulo 4 com alguns modelos de populações infinitas, de forma que adiaremos a discussão até lá.

A média amostral e a mediana amostral são influenciadas por valores fora da faixa de uma forma bastante diferente: muito para a média e nada para a mediana. Como o comportamento extremo dos dois valores é indesejável, consideraremos medidas alternativas que não sejam tão sensíveis quanto $\bar{x}$ e nem tão insensíveis como $\tilde{x}$. Para determinar essas alternativas, observe que $\bar{x}$ e $\tilde{x}$ são extremidades opostas da mesma "família" de medidas. Após o conjunto de dados ser ordenado, $\tilde{x}$ é calculado desprezando-se todos os valores possíveis em cada extremidade sem eliminar nada (deixando apenas um ou dois valores centrais) e obtendo a média do que restou. Por outro lado, para calcular $\bar{x}$, nada é desprezado antes de se obter a média. Para fazer uma comparação, a média envolve desprezar 0% de cada extremidade da amostra, enquanto, para a mediana, o máximo possível é desprezado de cada extremidade. Uma **média aparada** é algo intermediário entre $\bar{x}$ e $\tilde{x}$. Uma média aparada de 10%, por exemplo, seria calculada eliminando-se os 10% superiores e os 10% inferiores da amostra, obtendo-se, então, a média do restante.

Exemplo 1.15

Considere as 20 observações a seguir, ordenadas da menor para a maior, cada uma representando a vida útil (em horas) de um determinado tipo de lâmpada incandescente:

612	623	666	744	883	898	964	970	983	1003
1016	1022	1029	1058	1085	1088	1122	1135	1197	1201

A média das 20 observações é $\bar{x}$ = 965,0 e $\tilde{x}$ = 1009,5. A média aparada de 10% é obtida pela exclusão das duas menores observações (612 e 623) e as duas maiores (1197 e 1201) seguida do cálculo da média dos 16 valores restantes, para obter $\bar{x}_{tr(10)} = 979{,}1$. O efeito de truncar a média aqui é produzir um "valor central" ligeiramente acima da média ($\bar{x}$ é trazido para baixo por alguns poucos valores de vida útil) e ainda consideravelmente abaixo da mediana. De forma similar, a média aparada de 20% faz uma média dos 12 valores do meio para obter $\bar{x}_{tr(20)} = 999{,}9$, mais perto ainda da mediana. (Veja a Figura 1.17.)

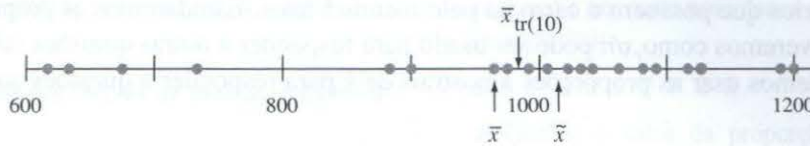


Figura 1.17 Gráfico de pontos de vida útil (em horas) de lâmpadas incandescentes ■

Geralmente, o uso da média aparada com proporção de aparagem moderada (entre 5% e 25%) produzirá uma medida que não é nem tão sensível a *outliers* como a média nem tão insensível quanto a mediana. Por esse motivo, as médias truncadas têm sido objeto de crescente atenção dos estatísticos para propósitos descritivos e inferenciais. Mais será dito sobre médias aparadas quando a estimativa por pontos for discutida no Capítulo 6. Finalmente, se a proporção de aparagem for representada por α e $n\alpha$ não for inteiro, não será óbvio como calcular a média aparada $100\alpha\%$. Por exemplo: se $\alpha = 0{,}10$ (10%) e $n = 22$, então $n\alpha = (22)(0{,}10) = 2{,}2$ e não é possível aparar 2,2 observações de cada extremidade da amostra ordenada. Nesse caso, a média aparada de 10% seria obtida primeiro com a retirada das duas observações de cada extremidade e pelo cálculo de $\bar{x}_{tr}$, seguida pela retirada de três observações de cada extremidade e pelo cálculo de $\bar{x}_{tr}$, e então pela interpolação dos dois valores para obtenção de $\bar{x}_{tr(10)}$.

Dados categorizados e proporção de amostras

Quando os dados são categorizados, uma distribuição de freqüência ou distribuição de freqüência relativa fornece um resumo tabular eficiente dos dados. Os indicadores numéricos naturais são, nessa situação, as freqüências individuais e as freqüências relativas. Por exemplo: se for feita uma pesquisa com indivíduos que possuem aparelhos de som para estudar a preferência de marca, cada indivíduo da amostra identificaria a marca do aparelho que possui. A partir disso poderíamos contar as pessoas que possuem aparelhos Sony, Pioneer, Marantz, entre outros. Considere a obtenção de uma amostra de uma população dicotômica, isto é, que consista em apenas duas categorias (votou ou não votou na eleição passada ou possui ou não um aparelho de som etc.). Se fizermos x representar o número da amostra na categoria 1, o número na categoria 2 será $n - x$. A freqüência relativa ou *proporção amostral* da categoria 1 será x/n e a proporção amostral da categoria 2 será $1 - x/n$. Vamos representar uma resposta da categoria 1 por 1 e uma resposta da categoria 2 por 0. Uma amostra de tamanho $n = 10$ pode então resultar em 1, 1, 0, 1, 1, 1, 0, 0, 1, 1. A média dessa amostra numérica é (já que o número de ocorrências do número $1 = x = 7$)

$$\frac{x_1 + \ldots + x_n}{n} = \frac{1 + 1 + 0 + \cdots + 1 + 1}{10} = \frac{7}{10} = \frac{x}{n} = \text{proporção amostral}$$

Esse resultado pode ser generalizado e resumido conforme segue: *Se em uma situação de dados categorizados focarmos a atenção em uma determinada categoria e codificarmos os resultados da amostra de forma que 1 seja registrado como um indivíduo da categoria e 0 para um indivíduo fora dela, a proporção amostral de indivíduos da categoria será a média amostral da seqüência de 1s e 0s.* Assim, uma média amostral pode ser usada para resumir os resultados de uma amostra categorizada. Essas observações também se aplicam a situações em que as categorias são definidas por valores agrupados em uma amostra ou população numérica (por exemplo: podemos querer saber se os indivíduos possuem seu automóvel atual há pelo menos cinco anos em vez de estudarmos o tempo exato de posse).

De forma análoga à proporção amostral x/n de indivíduos que estão em uma determinada categoria, representemos por p a proporção dos indivíduos da população inteira que pertencem à categoria. Como acontece com x/n, p é uma quantidade entre 0 e 1 e, enquanto x/n é uma característica da amostra, p é uma característica da população. A relação entre os dois é semelhante à relação entre $\tilde{x}$ e $\tilde{\mu}$ e entre $\bar{x}$ e μ. Em particular, usaremos x/n para fazer inferências sobre p. Se, por exemplo, uma amostra de 100 proprietários de carros revelar que 22 possuem seus carros há pelo menos 5 anos, podemos usar $22/100 = 0{,}22$ como uma estimativa pontual da proporção

de todos os proprietários que possuem o carro há pelo menos 5 anos. Estudaremos as propriedades de x/n como um estimador de p e veremos como x/n pode ser usado para responder a outras questões inferenciais. Com k categorias $(k > 2)$, podemos usar as proporções amostrais de k para responder a questões sobre as proporções da população $p_1, \ldots, p_k$.

Exercícios | Seção 1.3 (33–43)

33. O artigo "The Pedaling Technique of Elite Endurance Cyclists" *(Int. J. of Sport Biomechanics,* 1991, p. 29-53) relatou os dados a seguir sobre a potência de uma única perna de um ciclista em alta carga de trabalho:

244	191	160	187	180	176	174
205	211	183	211	180	194	200

a. Calcule e interprete a média e a mediana amostrais.
b. Suponha que a primeira observação tenha sido 204 em vez de 244. Como a média e a mediana seriam afetadas?
c. Calcule uma média aparada, eliminando a maior e a menor observações da amostra. Qual é a porcentagem de truncamento correspondente?
d. O artigo também relatou valores sobre a potência de uma única perna para uma carga de trabalho baixa. A média amostral de $n = 13$ observações foi $\bar{x} = 119{,}8$ (na verdade 119,7692) e a 14ª observação, um tipo de *outlier*, foi 159. Qual é o valor de $\bar{x}$ para a amostra toda?

34. Considere as observações a seguir sobre resistência ao cisalhamento (MPa) de uma junta soldada de uma determinada forma (de um gráfico no artigo "Diffusion of Silicon Nitride to Austenitic Stainless Steel without Interlayers," *Metallurgical Trans.*, 1993, p. 1835-1843):

22,2	40,4	16,4	73,7	36,6	109,9
30,0	4,4	33,1	66,7	81,5	

a. Determine o valor da média amostral.
b. Determine o valor da mediana amostral. Por que esse valor é tão diferente da média?
c. Calcule a média aparada, excluindo a menor e a maior observações. Qual é a porcentagem de truncamento correspondente? Como o valor de $\bar{x}_{tr}$ se compara à média e à mediana?

35. A pressão mínima de injeção (psi) em amostras de moldagem por injeção de milho de alta amilose foi determinada para oito amostras diferentes (pressões mais altas correspondem a maior dificuldade de processamento), resultando nas observações a seguir (de "Thermoplastic Starch Blends with a Polyethylene-Co-Vinyl Alcohol: Processability and Physical Properties," *Polymer Engr. and Science,* 1994, p. 17-23):

15,0 13,0 18,0 14,5 12,0 11,0 8,9 8,0

a. Determine os valores da média amostral, mediana amostral e média aparada de 12,5% e os compare.
b. Em quanto a menor observação da amostra, atualmente 8,0, pode ser aumentada sem afetar o valor da mediana amostral?
c. Suponha que desejemos que os valores da média e da mediana amostrais sejam expressos em quilogramas por polegada quadrada (ksi) em vez de psi. É necessário mudar as unidades de cada observação ou os valores calculados na parte (a) podem ser usados diretamente? *Dica:* 1kg = 2,2lb.

36. Vinte e seis trabalhadores de plataformas de petróleo *offshore* participaram de um exercício de fuga simulado, resultando nos dados a seguir (em segundos) para concluir a fuga ("Oxygen Consumption and Ventilation During Escape from an Offshore Platform," *Ergonomics,* 1997, p. 281-292):

389	356	359	363	375	424	325	394	402
373	373	370	364	366	364	325	339	393
392	369	374	359	356	403	334	397	

a. Construa um diagrama de caule e folha dos dados. Como ele sugere que a média e mediana serão comparadas?
b. Calcule os valores da média e da mediana amostrais. *Dica:* $\sum x_i = 9638$.
c. Em quanto o maior tempo, atualmente 424, pode ser aumentado sem afetar o valor da mediana amostral? Em quanto esse valor pode ser diminuído sem afetar o valor da mediana amostral?
d. Quais são os valores de $\bar{x}$ e $\tilde{x}$ quando as observações são reexpressas em minutos?

37. O artigo "Snow Cover and Temperature Relationships in North America and Eurasia" *(J. Climate and Applied Meteorology,* 1983, p. 460-469) usou técnicas estatísticas para relacionar a quantidade de cobertura de neve em cada continente com a temperatura média continental. Os dados apresentados incluíram as 10 observações a seguir sobre a cobertura de neve, em outubro, na Eurásia, durante 1970 e 1979 (em milhões de km²):

6,5 12,0 14,9 10,0 10,7 7,9 21,9 12,5 14,5 9,2

O que você descreveria como valor característico ou representativo da cobertura de neve em outubro para esse período e o que o levou a essa escolha?

38. Os valores de pressão sangüínea freqüentemente são informados com aproximação de 5 mmHg (100, 105, 110 etc.). Suponha que os valores reais (sem aproximação) de pressão sangüínea de nove indivíduos selecionados aleatoriamente sejam

118,6 127,4 138,4 130,0 113,7 122,0 108,3 131,5 133,2

a. Qual é a mediana dos valores de pressão sangüínea *informados*?

b. Suponha que a pressão sangüínea do segundo indivíduo seja 127,6 em vez de 127,4 (uma pequena alteração em um único valor). Como isso afeta a mediana dos valores informados? O que isso diz sobre a sensibilidade da mediana ao arredondamento ou agrupamento dos dados?

39. A propagação de trincas por fadiga em diversas peças de aeronaves tem sido objeto de muitos estudos nos últimos anos. Os dados a seguir consistem dos tempos de propagação (horas de vôo/10^4) para atingir um determinado tamanho de trinca em furos de fixadores propostos para uso em aeronaves militares ("Statistical Crack Propagation in Fastener Holes under Spectrum Loading," *J. Aircraft*, 1983, p. 1028-1032):

0,736 0,863 0,865 0,913 0,915 0,937 0,983 1,007
1,011 1,064 1,109 1,132 1,140 1,153 1,253 1,394

a. Calcule e compare os valores da média e da mediana amostrais.

b. Em quanto a maior observação da amostra pode ser diminuída sem afetar o valor da mediana?

40. Calcule a mediana amostral, a média aparada de 25%, a média aparada de 10% e a média amostral para os dados de concentração fornecidos no Exercício 27 e compare essas medidas.

41. Uma amostra de $n = 10$ automóveis foi selecionada e cada um deles foi sujeito a um teste de colisão a 5 mph. Representando um carro sem danos visíveis por S (de sucesso) e um carro com danos por F, os resultados são os seguintes:

S S F S S S F F S S

a. Qual é o valor da proporção amostral de sucessos x/n?

b. Substitua cada S por 1 e cada F por 0. Calcule então $\bar{x}$ para essa amostra codificada numericamente. Como $\bar{x}$ pode ser comparado a x/n?

c. Suponha que se decidiu incluir mais 15 carros no experimento. Quantos deles teriam de ser S para fornecer $x/n = 0{,}80$ para a amostra de 25 carros?

42. a. Se uma constante c é adicionada a cada x_i de uma amostra, resultando em $y_i = x_i + c$, como a média e a mediana amostrais dos y_i se relacionam com a média e a mediana dos x_i? Verifique suas hipóteses.

b. Se cada x_i é multiplicado por uma constante c, resultando em $y_i = cx_i$, responda à questão da parte (a). Verifique novamente suas hipóteses.

43. Um experimento para estudar a vida útil (em horas) de certo tipo de componente consiste em colocar dez componentes em operação e observá-los por 100 horas. Oito dos componentes apresentaram falhas nesse período e esses valores de vida útil foram registrados. Represente os tempos de vida útil dos dois componentes que ainda funcionam após 100 horas por 100+. As observações resultantes foram

48 79 100+ 35 92 86 57 100+ 17 29

Que medidas de tendência central discutidas nesta seção podem ser calculadas e quais são os valores dessas medidas? (*Nota:* Os dados deste experimento são "censurados pela direita".)

1.4 Medidas de Dispersão

Informar apenas a medida de tendência central fornece apenas informações parciais sobre um conjunto de dados ou uma distribuição. Diferentes amostras ou populações podem ter medidas de tendência central idênticas e apresentar diferenças entre si em outros aspectos importantes. A Figura 1.18 apresenta gráficos de pontos de três amostras com a mesma média e a mesma mediana, mas com dispersões diferentes ao redor do centro. A primeira amostra é a que apresenta maior dispersão, a terceira a menor e a segunda é intermediária em relação a elas.

Medidas de dispersão para dados amostrais

A medida de dispersão mais simples de uma amostra é a **amplitude**, a diferença entre o maior e o menor valores da amostra. Observe que o valor da amplitude da amostra 1 da Figura 1.18 é muito maior do que o da amostra 3, o que reflete maior dispersão na primeira amostra do que na terceira. Um defeito da amplitude, entretanto, é que ela depende apenas das duas observações mais extremas e não considera as posições dos $n - 2$ valores restantes. As amostras 1 e 2 na Figura 1.18 possuem amplitudes idênticas mas, se levarmos em conta as observações entre os dois extremos, há muito menos dispersão na segunda amostra do que na primeira.

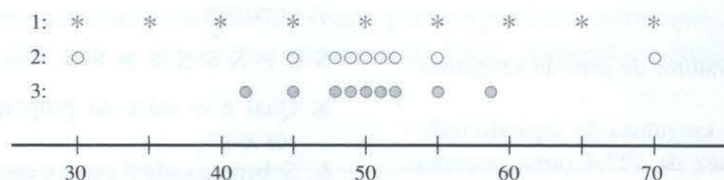


Figura 1.18 Amostras com medidas centrais idênticas mas com variabilidade diferente

Nossa principal medida de dispersão envolve os **desvios em relação à média,** $x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x}$. Ou seja, os desvios da média são obtidos pela subtração de $\bar{x}$ de cada uma das n observações da amostra. Um desvio será positivo se a observação for maior que a média (à direita da média no eixo das medidas) e negativo se a observação for menor que a média. Se todos os desvios forem pequenos em magnitude, todos os x_i estarão próximos à média e haverá pouca dispersão. Por outro lado, se alguns desvios forem grandes, alguns x_i estarão distantes de $\bar{x}$, indicando maior dispersão. Uma forma simples de combinar os desvios em uma única quantidade é calcular a sua média (somá-los e dividi-los por n). Infelizmente, há um problema grave com essa sugestão:

$$\text{somatória dos desvios} = \sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

de forma que o desvio será sempre zero. A demonstração usa diversas regras-padrão de somatória e o fato de que $\sum \bar{x} = \bar{x} + \bar{x} + \cdots + \bar{x} = n\bar{x}$:

$$\sum (x_i - \bar{x}) = \sum x_i - \sum \bar{x} = \sum x_i - n\bar{x} = \sum x_i - n\left(\frac{1}{n}\sum x_i\right) = 0$$

Como podemos alterar os desvios para quantidades não-negativas de forma que os desvios positivos e negativos não se cancelem quando forem somados? Uma possibilidade é trabalhar com os valores absolutos e calcular o desvio médio absoluto $\sum |x_i - \bar{x}| / n$. Como a operação em valor absoluto conduz a diversas dificuldades teóricas, considere os quadrados dos desvios $(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, \ldots, (x_n - \bar{x})^2$. Em vez de usar o quadrado do desvio médio $\sum (x_i - \bar{x})^2 / n$, por diversos motivos, dividiremos a soma dos quadrados dos desvios por $n - 1$ em vez de n.

DEFINIÇÃO

A **variância amostral**, representada por s^2, é dada por

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{S_{xx}}{n - 1}$$

O **desvio padrão amostral,** representado por s, é a raiz quadrada (positiva) da variância:

$$s = \sqrt{s^2}$$

A unidade de s é a mesma de cada um dos x_i. Se, por exemplo, as observações forem consumo de combustível em milhas por galão, podemos ter $s = 2{,}0$ milhas/galão. Uma interpretação grosseira do desvio padrão da amostra é que ele é o tamanho de um desvio típico ou representativo da média amostral para a amostra selecionada. Dessa forma, se $s = 2{,}0$ milhas/galão, então alguns x_i da amostra estão dentro do intervalo $\bar{x}$ mais de que 2,0, enquanto outros estão mais distantes; 2,0 é um desvio representativo (ou "padrão") do consumo médio de combustível. Se $s = 3{,}0$ para uma segunda amostra de carros de outro tipo, um desvio típico dessa amostra é cerca de 1,5 vez maior do que na primeira amostra, indicando maior dispersão da segunda.

Exemplo 1.16

A resistência é uma importante característica de materiais usados em casas pré-fabricadas. Cada um dos $n = 11$ elementos de placas pré-fabricadas foi submetido a um teste de tensão severo e a largura máxima (mm) das trincas resultantes foi registrada. Os dados fornecidos (Tabela 1.3) foram relatados no artigo "Prefabricated Ferrocement Ribbed Elements for Low-Cost Housing" (*J. Ferrocement,* 1984, p. 347-364). Conseqüências do arredondamento influenciam para a soma dos desvios não ser exatamente nula. O numerador de s^2 é 11,9359, portanto $s^2 = 11{,}9359/(11 - 1) = 11{,}9359/10 = 1{,}19359$ e $s = \sqrt{1{,}19359} = 1{,}0925$ mm. ■

Tabela 1.3 Dados do Exemplo 1.16

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
0,684	−0,9841	0,9685
2,540	0,8719	0,7602
0,924	−0,7441	0,5537
3,130	1,4619	2,1372
1,038	−,6301	0,3970
0,598	−1,0701	1,1451
0,483	−1,1851	1,4045
3,520	1,8519	3,4295
1,285	−0,3831	0,1468
2,650	0,9819	0,9641
1,497	−0,1711	0,0293
$\sum x_i = 18{,}349$	$\sum(x_i - \bar{x}) = -0{,}0001$	$S_{xx} = \sum(x_i - \bar{x})^2 = 11{,}9359$

$$\bar{x} = \frac{18{,}349}{11} = 1{,}6681$$

Dedução de s^2

Para explicar por que s^2 deve ser usado em vez do quadrado do desvio médio para medir a dispersão, observe primeiro que, enquanto s^2 mede a dispersão da amostra, há uma medida de dispersão da população denominada *variância da população.* Usaremos σ^2 (o quadrado da letra grega minúscula sigma) para representar a variância da população e σ para representar o desvio padrão da população (a raiz quadrada de σ^2). Quando a população é finita e consiste de N valores,

$$\sigma^2 = \sum_{i=1}^{N} (x_i - \mu)^2/N$$

que é a média de todos os quadrados dos desvios da média da população (no caso de população, o divisor é N e não $N - 1$). Definições mais gerais de σ^2 são apresentadas nos capítulos 3 e 4.

Da mesma forma que $\bar{x}$ é usado para fazer inferências sobre a média da população μ, devemos definir a variância da amostra de maneira que possa ser usada para fazer inferências sobre σ^2. Observe que σ^2 envolve os quadrados dos desvios em torno da média da população μ. Se conhecêssemos o valor real de μ, poderíamos então definir a variância da amostra como o quadrado do desvio médio da amostra x_i em torno de μ. Entretanto, o valor de μ quase nunca é conhecido, de modo que deve ser usada a soma dos quadrados dos desvios em torno de $\bar{x}$. Todavia, os x_i tendem a estar mais próximos de sua média do que da média da população μ, assim, para compensar tal fato, é usado o divisor n − 1 em vez de n. Em outras palavras, se usássemos o divisor n na fórmula da variância da amostra, quantidade resultante tenderia a subestimar σ^2 (gerar valores na média muito pequenos para a estimativa), enquanto a divisão pelo valor, ligeiramente menor, n − 1 corrige a subestimativa.

É costume se referir a s^2 como tendo $n - 1$ **graus de liberdade** (gl) como base. Essa terminologia resulta do fato de que, apesar de s^2 ter como base as n quantidades $x_1 - \bar{x}, x_2 - \bar{x}, \ldots x_n - \bar{x}$, sua soma é nula. Assim, especificar os valores de qualquer $n - 1$ das quantidades determina o valor restante. Por exemplo: se $n = 4$ e $x_1 - \bar{x} = 8$, $x_2 - \bar{x} = -6$ e $x_4 - \bar{x} = -4$, automaticamente $x_3 - \bar{x} = 2$, de forma que apenas três dos quatro valores de $x_i - \bar{x}$ serão determinados livremente (3 gl).

Uma fórmula para o cálculo de s^2

O cálculo dos quadrados dos desvios é tedioso, especialmente se for usada para $\bar{x}$ precisão decimal suficiente para prevenir os efeitos de arredondamento. Uma fórmula alternativa para o numerador de s^2 evita a necessidade de todas as subtrações para obter os desvios. A fórmula envolve $(\sum x_i)^2$, somando e depois obtendo os quadrados e $\sum x_i^2$, obtendo os quadrados e depois somando.

Uma expressão alternativa para o numerador de s^2 é

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

Demonstração Como $\bar{x} = \sum x_i/n$, $n\bar{x}^2 = (\sum x_i)^2/n$. Então,

$$\sum (x_i - \bar{x})^2 = \sum (x_i^2 - 2\bar{x} \cdot x_i + \bar{x}^2) = \sum x_i^2 - 2\bar{x}\sum x_i + \sum (\bar{x})^2$$
$$= \sum x_i^2 - 2\bar{x} \cdot n\bar{x} + n(\bar{x})^2 = \sum x_i^2 - n(\bar{x})^2$$

Exemplo 1.17

A quantidade de reflexão de luz pelas folhas foi usada para diversos propósitos, incluindo a avaliação da cor da plantação de grama, a estimativa do *status* de nitrogênio e a medida de biomassa. O artigo "Leaf Reflectance-Nitrogen-Chlorophyll Relations in Buffel-Grass" *(Photogrammetric Engr. and Remote Sensing*, 1985, p. 463-466) forneceu as observações a seguir, obtidas usando espectrofotogrametria, sobre a reflexão das folhas sob condições experimentais especificadas.

Observação	x_i	x_i^2	Observação	x_i	x_i^2
1	15,2	231,04	9	12,7	161,29
2	16,8	282,24	10	15,8	249,64
3	12,6	158,76	11	19,2	368,64
4	13,2	174,24	12	12,7	161,29
5	12,8	163,84	13	15,6	243,36
6	13,8	190,44	14	13,5	182,25
7	16,3	265,69	15	12,9	166,41
8	13,0	169,00			
				$\sum x_i = 216{,}1$	$\sum x_i^2 = 3168{,}13$

A fórmula de cálculo fornece

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 3168{,}13 - \frac{(216{,}1)^2}{15}$$
$$= 3168{,}13 - 3113{,}28 = 54{,}85$$

de onde $s^2 = S_{xx}/(n-1) = 54{,}85/14 = 3{,}92$ e $s = 1{,}98$. ■

O método alternativo pode resultar em valores de s^2 e s diferentes dos valores calculados, usando as definições. Essas diferenças se devem a efeitos de arredondamento e não serão importantes na maioria das amostras. Para minimizar os efeitos de arredondamento ao usar a fórmula alternativa, os cálculos intermediários devem ser feitos usando-se vários dígitos significativos a mais do que serão mantidos na resposta final. Como o numerador de s^2 é a soma de quantidades não-negativas (quadrados desvios), s^2 é com certeza não-negativo. Ainda assim, se for usada alternativa, particularmente com dados de pouca dispersão, um pequeno erro numérico pode resultar em um numerador negativo ($\sum x_i^2$ menor que $(\sum x_i)^2/n$). Se o valor de s^2 for negativo, foi cometido um erro de cálculo.

Diversas outras propriedades de s^2 podem facilitar esse cálculo.

PROPOSIÇÃO

Considere $x_1, x_2, \ldots, x_n$ como uma amostra e c como uma constante qualquer diferente de zero.

1. Se $y_1 = x_1 + c$, $y_2 = x_2 + c, \ldots, y_n = x_n + c$, então $s_y^2 = s_x^2$, e

2. Se $y_1 = cx_1, \ldots, y_n = cx_n$, então $s_y^2 = c^2 s_x^2$, $s_y = |c| s_x$,

onde s_x^2 é a variância amostral dos x e s_y^2 é a variância amostral dos y.

Em palavras, o resultado 1 informa que, se uma constante c for adicionada a (ou subtraída de) cada valor dos dados, a variância não será alterada. Isso é intuitivo, já que a adição ou a subtração de c altera a localização do conjunto de dados, mas mantém as distâncias entre os valores inalteradas. De acordo com o resultado 2, a multiplicação de cada x_i por c resulta em s^2, sendo multiplicado por um fator de c^2. Essas propriedades podem ser demonstradas pela observação de que $\bar{y} = \bar{x} + c$ no resultado 1 e de que $\bar{y} = c\bar{x}$ no resultado 2.

Boxplots

Diagramas de caule e folha e histogramas conduzem a impressões gerais sobre um conjunto de dados, enquanto um único valor como a média ou o desvio padrão enfoca apenas um aspecto dos dados. Nos últimos anos, um resumo esquemático denominado *boxplot* vem sendo usado para descrever as características mais proeminentes de conjuntos de dados. Essas características incluem (1) centro, (2) dispersão, (3) a extensão e a natureza de qualquer desvio em relação à simetria e (4) a identificação de *outliers*, observações que normalmente estão distantes da maior parte dos dados. Como apenas um *outlier* pode afetar drasticamente os valores de $\bar{x}$ e s, um *boxplot* é baseado em medidas "resistentes" à presença de alguns *outliers*: a mediana e uma medida de dispersão denominada *dispersão entre os quartos*.

DEFINIÇÃO

Ordene as n observações da menor para a maior e então separe a metade menor da maior. A mediana $\tilde{x}$ estará incluída em ambas as partes se n for ímpar. Então o **quarto inferior** será a mediana da metade menor e o **quarto superior** será a mediana da metade maior. Uma medida de dispersão resistente a *outliers* é a **dispersão entre os quartos** f_s, dada por

$$f_s = \text{quarto superior} - \text{quarto inferior}$$

Grosso modo, a dispersão entre os quartos não será alterada pelas posições das observações nos menores 25% ou nos maiores 25% dos dados.

O *boxplot* mais simples tem base no seguinte resumo de cinco números:

menor x_i quarto inferior mediana quarto superior maior x_i

Primeiro, desenhe um eixo de medida horizontal. Então, coloque um retângulo sobre o eixo; a extremidade esquerda do retângulo estará no quarto inferior e a extremidade direita estará no quarto superior (de forma que a largura da caixa = f_s). Trace um segmento de reta vertical ou outro símbolo dentro do retângulo na posição da mediana. A posição do símbolo da mediana em relação às duas extremidades indica informações sobre o desvio nos 50% centrais dos dados. Por fim, desenhe "bigodes" saindo de cada extremidade do retângulo para as observações maior e menor. Um *boxplot* com uma orientação vertical também pode ser desenhado, fazendo-se modificações óbvias no processo de construção.

Exemplo 1.18

O ultra-som foi usado para obter informações sobre dados de corrosão na espessura da chapa do assoalho de um reservatório elevado usado para armazenar óleo bruto ("Statistical Analysis of UT Corrosion Data from Floor Plates of a Crude Oil Aboveground Storage Tank," *Materials Eval.*, 1994, p. 846-849). Cada observação é a maior profundidade do orifício na placa, expressa em milipolegadas.

40 52 55 60 70 75 85 85 90 90 92 94 94 95 98 100 115 125 125

O resumo de cinco números segue:

menor x_i = 40 quarto inferior = 72,5 $\tilde{x}$ = 90 quarto superior = 96,5
maior x_i = 125

A Figura 1.19 exibe o *boxplot* resultante. A extremidade direita da caixa está muito mais próxima da mediana do que a esquerda, indicando um desvio substancial na parte central dos dados. A largura da caixa (f_s) também é relativamente grande em relação à amplitude dos dados (distância entre as pontas dos bigodes).

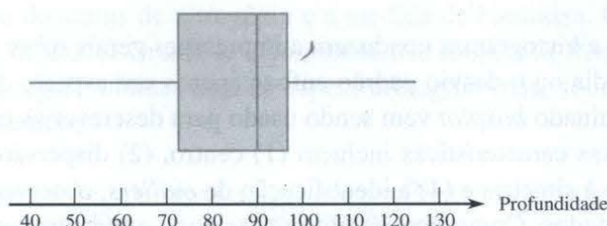


Figura 1.19 Um *boxplot* dos dados de corrosão

A Figura 1.20 mostra o resultado do MINITAB de uma solicitação de descrição dos dados de corrosão. A média aparada é a média das 17 observações que permanecem após a exclusão do maior e do menor valores (porcentagem de truncamento ≈ 5%). Q1 e Q3 são os quartis inferior e superior. Eles são similares aos quartos, mas calculados de uma forma ligeiramente diferente. A Média SE é $s/\sqrt{n}$, quantidade que será importante no trabalho subseqüente em inferências de μ.

```
Variável          N     Média     Mediana    Média Ap   Desv Padrão   Média SE
profundidade     19     86,32       90,00       86,76         23,32       5,35

Variável     Mínimo    Máximo          Q1          Q3
profundidade  40,00    125,00       70,00       98,00
```

Figura 1.20 Descrição em MINITAB dos dados de profundidade do orifício ■

Boxplots com Outliers

Um *boxplot* pode ser formatado para indicar explicitamente a presença de *outliers*.

DEFINIÇÃO Qualquer observação distante mais do que $1{,}5f_s$ do quarto mais próximo é um ***outlier***. Um *outlier* é **extremo** se estiver a mais de $3f_s$ do quarto mais próximo. Caso contrário, é **moderado**.

Muitos procedimentos inferenciais se baseiam na suposição de que a amostra é proveniente de uma distribuição normal. Mesmo um único *outlier* extremo na amostra avisa o investigador de que tais procedimentos não devem ser usados e a presença de diversos *outliers* moderados passa a mesma mensagem.

Modifiquemos agora nossa construção anterior de um *boxplot*, desenhando um "bigode" em cada extremidade da "caixa", para que a menor e a maior observações *não* sejam *outliers*. Cada *outlier* moderado é representado por um círculo cheio e cada *outlier* extremo é representado por um círculo vazio. Alguns softwares estatísticos não fazem distinções entre *outliers* extremos e moderados.

Exemplo 1.19

Os efeitos de descargas parciais na degradação de materiais de cavidades isolante têm importantes implicações na vida útil de componentes de alta voltagem. Consideremos a seguinte amostra de $n = 25$ larguras de pulso de descargas lentas em uma cavidade cilíndrica de polietileno. (Esses dados são consistentes com um histograma de 250 observações no artigo "Assessment of Dielectric Degradation by Ultrawide-band PD Detection," *IEEE Trans. on Dielectrics and Elec. Insul.*, 1995, p. 744-760.) O autor do artigo nota o impacto de diversas ferramentas estatísticas na interpretação dos dados de descarga.

5,3	8,2	13,8	74,1	85,3	88,0	90,2	91,5	92,4	92,9	93,6	94,3	94,8
94,9	95,5	95,8	95,9	96,6	96,7	98,1	99,0	101,4	103,7	106,0	113,5	

Os indicadores relevantes são

$\tilde{x} = 94{,}8$ quarto inferior $= 90{,}2$ quarto superior $= 96{,}7$

$f_s = 6{,}5$ $1{,}5f_s = 9{,}75$ $3f_s = 19{,}50$

Dessa forma, qualquer observação menor que $90{,}2 - 9{,}75 = 80{,}45$ ou maior que $96{,}7 + 9{,}75 = 106{,}45$ é um *outlier*. Há um *outlier* na extremidade superior da amostra e quatro *outliers* na inferior. Como $90{,}2 - 19{,}5 = 70{,}7$, as três observações: 5,3, 8,2 e 13,8 são *outliers* extremos. Os outros dois *outliers* são moderados. Os "bigodes" se estendem até 85,3 e 106,0, as observações mais extremas que não são *outliers*. O *boxplot* resultante está na Figura 1.21. Há desvio negativo grande na parte central da amostra, assim como na amostra inteira.

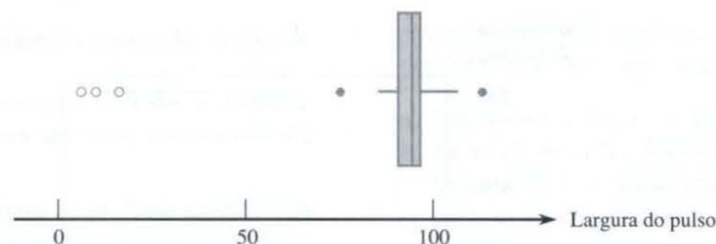


Figura 1.21 Um *boxplot* dos dados de largura de pulso que exibem *outliers* extremos e moderados ■

Boxplots Comparativos

Um *boxplot* comparativo ou lado a lado é uma forma muito eficiente de revelar semelhanças e diferenças entre dois ou mais conjuntos de dados consistindo de observações da mesma variável.

Exemplo 1.20

Nos últimos anos, algumas evidências sugerem que a alta concentração de radônio em ambientes fechados pode estar ligada ao desenvolvimento de cânceres infantis, mas muitos profissionais da saúde não estão convencidos. Um artigo recente ("Indoor Radon and Childhood Cancer", *The Lancet*, 1991, p. 1537-1538) apresentou os seguintes dados sobre a concentração de radônio (Bq/m^3) em duas amostras diferentes de casas. A primeira amostra consistia de casas em que havia residido uma criança com diagnóstico de câncer. As casas da segunda amostra não possuíam casos registrados de câncer infantil. A Figura 1.22 apresenta um diagrama de caule e folha dos dedos.

1. Câncer		2. Ausência de câncer
9683795	0	95768397678993
86071815066815233150	1	12271713114
12302731	2	99494191
8349	3	839
5	4	
7	5	55
	6	
	7	Caule: Dígito das dezenas
HI: 210	8	5 Folha: Dígito das unidades

Figura 1.22 Diagrama de caule e folha do Exemplo 1.20

As quantidades numéricas resumidas são as seguintes:

	$\bar{x}$	$\tilde{x}$	s	f_s
Câncer	22,8	16,0	31,7	11,0
Sem câncer	19,2	12,0	17,0	18,0

Os valores da média e da mediana sugerem que a amostra de câncer está centrada um pouco à direita da amostra de ausência de câncer na escala de medida. A média, entretanto, exagera a grandeza do desvio, principalmente devido ao valor 210 na amostra de câncer. Os valores de s sugerem dispersão maior na amostra de câncer do que na amostra de ausência de câncer, mas essa impressão é contrariada pela quarta dispersão. Novamente, a observação 210, um *outlier* extremo, é a culpada. A Figura 1.23 mostra um *boxplot* comparativo do pacote de software S-Plus. A caixa dos indivíduos com ausência de câncer é mais extensa se comparada à caixa de câncer ($f_s = 18$ *vs.* $f_s = 11$), além disso, as posições das retas das medianas mostram desvio maior na metade da amostra de ausência de câncer do que na amostra de câncer. Os *outliers* são representados por segmentos de reta horizontais e não há distinção entre *outliers* moderados e extremos.

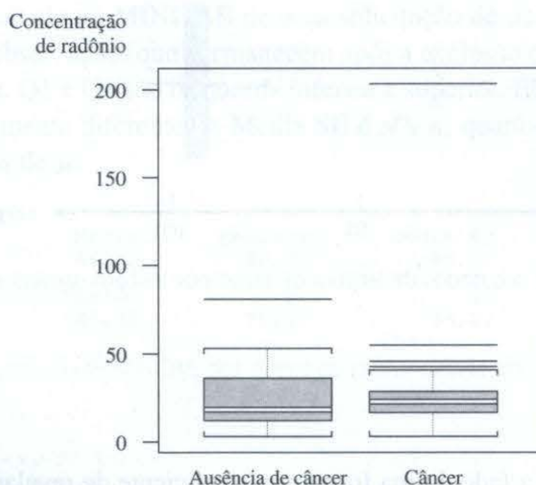


Figura 1.23 Um *boxplot* dos dados do Exemplo 1.20, no S-Plus ■

Exercícios Seção 1.4 (44–61)

44. O artigo "Oxygen Consumption During Fire Suppression: Error of Heart Rate Estimation" *(Ergonomics,* 1991, p. 1469-1474) informou os dados a seguir sobre consumo de oxigênio (mL/kg/min) para uma amostra de 10 bombeiros em uma simulação de supressão de incêndio:

29,5 49,3 30,6 28,2 28,0 26,3 33,9 29,4 23,5 31,6

Calcule:

a. A amplitude amostral.

b. A variância amostral s^2 pela definição (ou seja, primeiro calculando os desvios e depois obtendo os quadrados etc.).

c. O desvio padrão amostral.

d. s^2 usando o método alternativo.

45. O valor do módulo de Young (GPa) foi determinado para chapas fundidas feitas de algumas substâncias metálicas, resultando nas observações a seguir ("Strength and Modulus of a Molybdenum-Coated Ti-25A1-10Nb-3U-lMo Intermetallic," *J. of Materials Engr. and Performance,* 1997, p. 46-50):

116,4 115,9 114,6 115,2 115,8

a. Calcule $\bar{x}$ e os desvios em relação à média.

b. Use os desvios calculados na parte (a) para obter a variância amostral e o desvio padrão amostral.

c. Calcule s^2 usando a fórmula do numerador S_{xx}.

d. Subtraia 100 de cada observação para obter uma amostra de valores transformados. Agora calcule a variância amostral desses valores transformados e a compare ao s^2 dos dados originais.

46. As observações a seguir da viscosidade estabilizada (cP) para amostras de certo tipo de asfalto com 18% de borracha adicionada são do artigo "Viscosity Characteristics of Rubber-Modified Asphalts" *(J. of Materials in Civil Engr.,* 1996, p. 153-156):

2781 2900 3013 2856 2888

a. Quais são os valores da média amostral e da mediana amostral?

b. Calcule a variância amostral usando a fórmula. *(Dica:* Primeiro subtraia um valor conveniente de cada observação.)

47. Calcule e interprete os valores da mediana amostral, da média amostral e do desvio padrão amostral das observações a seguir da resistência à ruptura (MPa, lidas de um gráfico de "Heat-Resistant Active Brazing of Silicon Nitride: Mechanical Evaluation of Braze Joints," *Welding J.,* August, 1997):

87 93 96 98 105 114 128 131 142 168

48. O Exercício 36 na Seção 1.3 apresentou uma amostra de 26 tempos de fuga dos trabalhadores de uma plataforma de petróleo em uma simulação de fuga. Calcule e interprete o desvio padrão amostral. *(Dica:* $\sum x_i = 9638$ e $\sum x_i^2 = 3.587.566$.)

49. Um estudo da relação entre idade e diversas funções visuais (como precisão e percepção de profundidade) informou as seguintes observações da área de lâmina escleral (mm^2) nas extremidades do nervo óptico humano ("Morphometry of Nerve Fiber Bundle Pores in the Optic Nerve Head of the Human," *Experimental Eye Research,* 1988, p. 559-568):

2,75 2,62 2,74 3,85 2,34 2,74 3,93 4,21 3,88
4,33 3,46 4,52 2,43 3,65 2,78 3,56 3,01

a. Calcule $\sum x_i$ e $\sum x_i^2$.

b. Use os valores calculados na parte (a) para obter a variância amostral s^2 e o desvio padrão amostral s.

50. Em 1997, uma mulher processou um fabricante de teclados de computadores, sob a acusação de lesões por esforços repetitivos causados pelo teclado *(Genessy v. Digital Equipment Corp.).* O pleito era de cerca de 3,5 milhões de dólares por danos físicos, mas a corte negou esse valor pois julgou a indenização exagerada. Ao fazer essa determinação, a corte identificou um grupo "normativo" de 27 casos similares e especificou como razoável uma indenização limitada por dois desvios padrão em relação à média das indenizações dos 27 casos. As 27 indenizações foram (em milhares de dólares) 37, 60, 75, 115, 135, 140, 149, 150, 238, 290, 340, 410, 600, 750, 750, 750, 1050, 1100, 1139, 1150, 1200, 1200, 1250, 1576, 1700, 1825 e 2000, das quais $\sum x_i = 20.179$, $\sum x_i^2 = 24.657.511$. Qual é o valor máximo que pode ser indenizado pela regra de dois desvios padrão?

51. O artigo "A Thin-Film Oxygen Uptake Test for the Evaluation of Automotive Crankcase Lubricants" *(Lubric. Engr.,* 1984, p. 75-83) informou os seguintes dados sobre tempo de oxidação-indução (min) para diversos óleos comerciais:

87 103 130 160 180 195 132 145 211 105 145
153 152 138 87 99 93 119 129

a. Calcule a variância e o desvio padrão amostrais.

b. Se as observações fossem especificadas em horas, quais seriam os valores resultantes para a variância e para o desvio padrão amostrais? Responda sem reescrever os valores.

52. Os primeiros quatro desvios em relação à média de uma amostra de $n = 5$ tempos de reação foram 0,3, 0,9, 1,0 e 1,3. Qual é o quinto desvio em relação à média? Forneça uma amostra para a qual esses são os cinco desvios em relação à média.

53. Reconsidere os dados sobre a área de lâmina escleral do Exercício 49.

a. Determine os quartos inferior e superior.

b. Calcule o valor da dispersão entre os quartos.

c. Se os dois maiores valores da amostra, 4,33 e 4,52, fossem 5,33 e 5,52, como f_s seria afetado? Explique.

d. Em quanto a observação 2,34 pode ser aumentada sem afetar f_s? Explique.

e. Se uma 18ª observação, x_{18} = 4,60, fosse adicionada à amostra, qual seria o valor de f_s?

54. Reconsidere as observações de resistência ao cisalhamento (MPa) apresentadas no Exercício 34 deste capítulo:

22,2	40,4	16,4	73,7	36,6	109,9
30,0	4,4	33,1	66,7	81,5	

a. Quais são os valores dos quartos e qual é o valor de f_s?

b. Construa um *boxplot* com base no resumo de cinco valores e comente suas características.

c. Quão grande ou pequena deve ser uma observação para se qualificar como um *outlier*? E como um *outlier* extremo?

d. Em quanto a maior observação pode ser diminuída sem afetar f_s?

55. Segue um diagrama de caule e folha dos dados de tempos de fuga apresentados no Exercício 36 deste capítulo.

32	55
33	49
34	
35	6699
36	34469
37	03345
38	9
39	2347
40	23
41	
42	4

a. Calcule o valor da dispersão entre os quartos.

b. Há algum *outlier* na amostra? Algum *outlier* extremo?

c. Construa um *boxplot* e comente suas características.

d. Em quanto a maior observação, 424, pode ser diminuída sem afetar o valor da dispersão entre os quartos?

56. A quantidade de contaminação por alumínio (ppm) em certo tipo de plástico foi determinada para uma amostra de 26 espécimes de plástico, resultando nos dados a seguir ("The Lognormal Distribution for Modeling Quality Data when the Mean Is Near Zero," *J. of Quality Technology*, 1990, p. 105-110):

30	30	60	63	70	79	87	90	101
102	115	118	119	119	120	125	140	145
172	182	183	191	222	244	291	511	

Construa um *boxplot* que mostre *outliers* e comente suas características.

57. Uma amostra de 20 garrafas de certo tipo de vidro foi selecionada e a resistência à pressão interna de cada garrafa foi determinada. Considere as seguintes informações parciais da amostra:

mediana = 202,2 quarto inferior = 196,0
quarto superior = 216,8

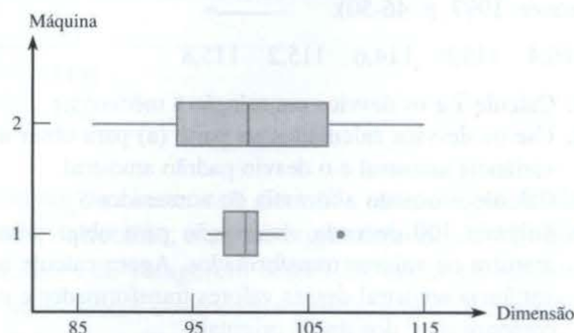
Três observações menores 125,8 188,1 193,7
Três observações maiores 221,3 230,5 250,2

a. Há algum *outlier* na amostra? Algum *outlier* extremo?

b. Construa um *boxplot* que mostre *outliers* e comente suas características.

58. Uma empresa usa duas máquinas diferentes para fabricar certo tipo de peça. Durante um turno, uma amostra de n = 20 peças produzidas por cada máquina é selecionada e o valor de uma importante dimensão de cada peça é determinado. O *boxplot* comparativo da figura a seguir foi construído a partir dos dados resultantes. Compare e destaque as diferenças entre as duas amostras.

Boxplot comparativo do Exercício 58



59. A concentração de cocaína no sangue (mg/L) foi determinada para uma amostra de indivíduos que morreram de delírio induzido por cocaína (ED) e para uma amostra de indivíduos que morreram de overdose de cocaína sem delírio. O tempo de sobrevida das pessoas em ambos os grupos foi de, no máximo, 6 horas. Os dados a seguir foram obtidos de um *boxplot* comparativo do artigo "Fatal Excited Delirium Following Cocaine Use" (*J. of Forensic Sciences*, 1997, p. 25-31).

ED 0 0 0 0 0,1 0,1 0,1 0,1 0,2 0,2 0,3 0,3
0,3 0,4 0,5 0,7 0,8 1,0 1,5 2,7 2,8
3,5 4,0 8,9 9,2 11,7 21,0

Não-ED 0 0 0 0 0,1 0,1 0,1 0,1 0,2 0,2 0,2
0,3 0,3 0,3 0,4 0,5 0,5 0,6 0,8 0,9 1,0
1,2 1,4 1,5 1,7 2,0 3,2 3,5 4,1
4,3 4,8 5,0 5,6 5,9 6,0 6,4 7,9
8,3 8,7 9,1 9,6 9,9 11,0 11,5
12,2 12,7 14,0 16,6 17,8

a. Determine as medianas, quartos e quartas dispersões das duas amostras.

b. Há algum *outlier* nas amostras? Algum *outlier* extremo?

c. Construa um *boxplot* comparativo e use-o como base para comparar e destacar as diferenças das amostras ED e não-ED.

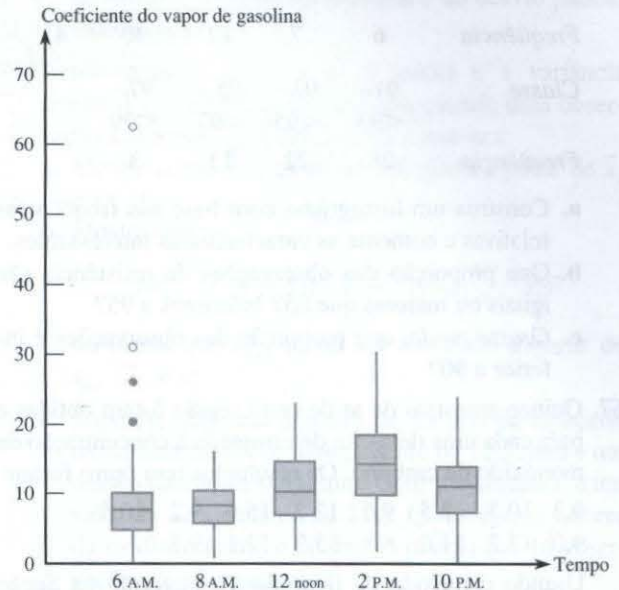
60. Foram obtidas observações sobre a resistência à explosão (lb/in^2) de soldas de fechamento de bocais de teste e soldas de bocais de canísteres de produção ("Proper Procedures Are the Key to Welding Radioactive Waste Cannisters," *Welding J.*, Aug. 1997, p. 61-67).

Teste	7200	6100	7300	7300	8000	7400
	7300	7300	8000	6700	8300	
Canister	5250	5625	5900	5900	5700	6050
	5800	6000	5875	6100	5850	6600

Construa um *boxplot* comparativo e faça comentários sobre os aspectos interessantes (o artigo citado não inclui a figura, mas os autores comentaram que ela foi vista).

61. O seguinte *boxplot* comparativo sobre coeficientes de vapor de gasolina para veículos em Detroit foi exibido no artigo "Receptor Modeling Approach to VOC Emission Inventory Validation" (*J. of Envir. Engr.,* 1995, p. 483-490). Comente as características interessantes.

Boxplot comparativo do Exercício 61



Exercícios Suplementares (62–83)

62. Considere as seguintes informações do limite de resistência à tração (lb/in) de uma amostra de $n = 4$ espécimes de fio rígido de cobre e zircônio (de "Characterization Methods for Fine Copper Wire" *Wire J. Intl.*, ago., 1997, p. 74-80):

$\bar{x} = 76831$ $s = 180$ menor $x_i = 76683$
maior $x_i = 77048$

Determine os valores das duas observações da metade da amostra (e não o faça adivinhando!).

63. Três valores diferentes de vazão C_2F_6 (SCCM) foram consideradas em um experimento para investigar o efeito da vazão na uniformidade (%) da extremidade de um *wafer* de silício usado na fabricação de circuitos integrados, resultando nos dados a seguir:

Vazão						
125	2,6	2,7	3,0	3,2	3,8	4,6
160	3,6	4,2	4,2	4,6	4,9	5,0
200	2,9	3,4	3,5	4,1	4,6	5,1

Compare e destaque as observações de uniformidade resultantes dessas diferentes vazões.

64. A quantidade de radiação recebida em uma estufa tem importante papel na determinação da taxa de fotossíntese. As observações a seguir sobre radiação solar foram lidas de um gráfico no artigo "Radiation Components over Bare and Planted Soils in a Greenhouse" (*Solar Energy,* 1990, p. 1011-1016).

6,3	6,4	7,7	8,4	8,5	8,8	8,9
9,0	9,1	10,0	10,1	10,2	10,6	10,6
10,7	10,7	10,8	10,9	11,1	11,2	11,2
11,4	11,9	11,9	12,2	13,1		

Use alguns dos métodos discutidos neste capítulo para descrever e resumir os dados.

65. Os seguintes dados sobre emissões de HC e CO para um determinado veículo foram fornecidos na introdução do capítulo.

HC (gm/mi)	13,8	18,3	32,2	32,5
CO (gm/mi)	118	149	232	236

a. Calcule os desvios padrão amostrais das observações de HC e CO. A crença parece ter fundamento?

b. O *coeficiente de dispersão amostral* $s/\bar{x}$ *(ou 100* $s/\bar{x}$) avalia o valor da dispersão em relação à média. Valores desses coeficientes de diversos conjuntos de dados podem ser comparados para determinar quais conjuntos de dados exibem mais ou menos dispersão. Faça uma comparação dos dados fornecidos.

66. A distribuição de freqüência a seguir, de observações sobre resistência à ruptura (MPa) de barras cerâmicas tratadas em um determinado forno, foi explicada no artigo "Evaluating Tunnel Kiln Performance" (*Amer. Ceramic Soc. Bull.,* ago. 1997, p. 59-63).

Classe	81–<83	83–<85	85–<87	87–<89	89–<91
Freqüência	6	7	17	30	43

Classe	91–<93	93–<95	95–<97	97–<99
Freqüência	28	22	13	3

a. Construa um histograma com base nas freqüências relativas e comente as características interessantes.
b. Que proporção das observações de resistência são iguais ou maiores que 85? Inferiores a 95?
c. *Grosso modo*, que proporção das observações é inferior a 90?

67. Quinze amostras de ar de certa região foram obtidas e para cada uma delas foi determinada a concentração de monóxido de carbono. Os resultados (em ppm) foram:

9,3 10,7 8,5 9,6 12,2 15,6 9,2 10,5
9,0 13,2 11,0 8,8 13,7 12,1 9,8

Usando o método de interpolação sugerido na Seção 1.3, calcule a média aparada 10%.

68. a. Para que valor de c a quantidade $\sum(x_i - c)^2$ é minimizada? *(Dica:* Obtenha a derivada em relação a c, iguale a 0 e resolva a equação).
b. Usando o resultado da parte (a), qual das duas quantidades $\sum(x_i - \bar{x})^2$ e $\sum(x_i - \mu)^2$ será menor que a outra (assumindo que $\bar{x} \neq \mu$)?

69. a. Considere a e b como constantes e $y_i = ax_i + b$ para $i = 1, 2, \ldots, n$. Quais são as relações entre $\bar{x}$ e $\bar{y}$ e entre s_x^2 e s_y^2?
b. Uma amostra de temperaturas iniciais de uma determinada reação química resultou em uma média amostral (°C) de 87,3 e um desvio padrão amostral de 1,04. Quais são a média e o desvio padrão em °F? *(Dica:* $F = \frac{9}{5}C + 32$.)

70. O elevado consumo de energia durante o exercício continua após o fim do treino. Como as calorias queimadas após o exercício contribuem para a perda de peso e têm outras conseqüências, é importante entender esse processo. O estudo "Effect of Weight Training Exercise and Treadmill Exercise on Post-Exercise Oxygen Consumption" *(Medicine and Science in Sports and Exercise,* 1998, p. 518-522*)* relatou os dados a seguir de um estudo em que o consumo de oxigênio (litros) foi medido continuamente por 30 minutos para cada um dos 15 indivíduos após um exercício de levantamento de peso e após um exercício em esteira rolante.

Indivíduo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Levantamento (x) *de peso*	14,6	14,4	19,5	24,3	16,3	22,1	23,0	18,7	19,0	17,0	19,1	19,6	23,2	18,5	15,9
Esteira (y) *rolante*	11,3	5,3	9,1	15,2	10,1	19,6	20,8	10,3	10,3	2,6	16,6	22,4	23,6	12,6	4,4

a. Construa um *boxplot* comparativo das observações de levantamento de peso e de esteira rolante e comente as características encontradas.
b. Como os dados estão na forma de pares (x, y), com as medidas x e y da mesma variável em duas condições diferentes, é natural enfocar as diferenças entre os pares: $d_1 = x_1 - y_1, \ldots, d_n = x_n - y_n$. Construa um *boxplot* das diferenças das amostras. O que ele sugere?

71. Segue uma descrição do MINITAB dos dados de resistência fornecidos no Exercício 13.

```
Variável         N  Média Mediana Média Ap DesvPad Média SE
resistência    153 135,39  135,40   135,41    4,59     0,37

Variável    Mínimo Máximo      Q1       Q3
resistência 122,20 147,70  132,95   138,25
```

a. Comente as características interessantes (os quartis e quartos são praticamente idênticos aqui).
b. Construa um *boxplot* dos dados com base nos quartis e comente suas características.

72. Distúrbios causados por ansiedade e seus sintomas normalmente podem ser tratados com medicamentos à base de benzodiazepina. É sabido que animais expostos a estresse apresentam redução da capacidade de absorção de receptores de benzodiazepina no córtex frontal. O artigo "Decreased Benzodiazepine Receptor Binding in Prefrontal Cortex in Combat-Related Posttraumatic Stress Disorder" *(Amer. J. of Psychiatry,* 2000, p. 1120-1126) descreveu o primeiro estudo da redução da capacidade de absorção de receptores de benzodiazepina em indivíduos que sofriam de DSPT. Os dados a seguir, relacionados a uma medida da capacidade de absorção (distribuição ajustada do volume), foram tirados de um gráfico do artigo.

DSPT: 10, 20, 25, 28, 31, 35, 37, 38, 38, 39, 39, 42, 46

Saudável: 23, 39, 40, 41, 43, 47, 51, 58, 63, 66, 67, 69, 72

Use os diversos métodos deste capítulo para descrever e resumir os dados.

73. O artigo "Can We Really Walk Straight?" *(Amer. J. of Physical Anthropology,* 1992, p. 19-27*)* relatou um experimento em que cada um de 20 homens saudáveis foi solicitado a caminhar da forma mais reta possível até um alvo a 60m de distância com velocidade normal. Considere as observações a seguir sobre a cadência (número de passos por segundo):

0,95 0,85 0,92 0,95 0,93 0,86 1,00 0,92 0,85 0,81
0,78 0,93 0,93 1,05 0,93 1,06 1,06 0,96 0,81 0,96

Use os métodos desenvolvidos neste capítulo para resumir os dados e inclua interpretações ou discussões, se apropriado. *(Nota:* O autor do artigo usou uma análise estatística sofisticada para concluir que as pessoas não podem andar em linha reta e sugeriu diversas explicações para isso.)

74. A **moda** de um conjunto de dados numéricos é o valor que ocorre mais freqüentemente no conjunto.

a. Determine a moda dos dados de cadência fornecidos no Exercício 73.

b. Para uma amostra categorizada, como você definiria a categoria modal?

75. Foram selecionadas amostras de três tipos de corda e o limite de fadiga (MPa) foi determinado para cada amostra, resultando os dados a seguir.

Tipo 1	350	350	350	358	370	370	370	371
	371	372	372	384	391	391	392	
Tipo 2	350	354	359	363	365	368	369	371
	373	374	376	380	383	388	392	
Tipo 3	350	361	362	364	364	365	366	371
	377	377	377	379	380	380	392	

a. Construa um boxplot comparativo e comente as semelhanças e diferenças.

b. Construa um gráfico de pontos (*dotplot*) comparativo (um *dotplot* para cada amostra com uma escala comum). Comente as semelhanças e diferenças.

c. O boxplot comparativo da parte (a) fornece uma avaliação informativa das semelhanças e diferenças? Explique seu raciocínio.

76. As três medidas de tendência central apresentadas neste capítulo são a média, a mediana e a média aparada. Duas medidas de tendência central adicionais, usadas ocasionalmente, são o centro da amplitude, que é a média entre a menor e a maior observações e a média dos quartos, que é a média dos dois quartos. Quais destas cinco medidas de tendência central são indiferentes aos efeitos de *outliers* e quais não são? Explique seu raciocínio.

77. Considere os dados a seguir sobre tempo de conserto ativo (horas) para uma amostra de $n = 46$ rádios de comunicação aéreos:

0,2	0,3	0,5	0,5	0,5	0,6	0,6	0,7	0,7	0,7	0,8	0,8
0,8	1,0	1,0	1,0	1,0	1,1	1,3	1,5	1,5	1,5	1,5	2,0
2,0	2,2	2,5	2,7	3,0	3,0	3,3	3,3	4,0	4,0	4,5	4,7
5,0	5,4	5,4	7,0	7,5	8,8	9,0	10,3	22,0	24,5		

Construa:

a. Um diagrama de caule e folha em que os dois maiores valores são exibidos separadamente em uma linha denominada HI;

b. Um histograma com base em seis intervalos de classe com 0 como limite inferior do primeiro intervalo e as larguras de intervalo de 2, 2, 2, 4, 10 e 10, respectivamente.

78. Considere uma amostra $x_1, x_2, \ldots, x_n$ e suponha que os valores de $\bar{x}$, s^2 e s tenham sido calculados.

a. Assuma $y_i = x_i - \bar{x}$ para $i = 1, \ldots, n$. Como os valores de s^2 e s dos y_i se comparam aos valores correspondentes dos x_i? Explique.

b. Assuma $z_i = (x_i - \bar{x})/s$ para $i = 1, \ldots, n$. Quais são os valores da variância amostral e do desvio padrão amostral dos z_i?

79. Representemos por $\bar{x}_n$ e s_n^2 a média e a variância amostral $\bar{x}_{n+1}$ e s_{n+1}^2, as quantidades quando uma observação adicional x_{n+1} é adicionada à amostra:

a. Mostre como $\bar{x}_{n+1}$ pode ser calculado a partir de $\bar{x}_n$ e x_{n+1}.

b. Mostre que

$$ns_{n+1}^2 = (n-1)s_n^2 + \frac{n}{n+1}(x_{n+1} - \bar{x}_n)^2$$

de forma que s_{n+1}^2 possa ser calculado a partir de x_{n+1}, $\bar{x}_n$, e s_n^2.

c. Suponha que uma amostra de 15 fios de tapeçaria tenha uma média de alongamento de 12,58mm e um desvio padrão de 0,512mm. O 16º fio resulta em um valor de alongamento de 11,8. Quais são os valores da média amostral e do desvio padrão das 16 observações de alongamento?

80. As distâncias das rotas de ônibus em qualquer sistema de trânsito normalmente variam de um trajeto para outro. O artigo "Planning of City Bus Routes" (*J. of the Institution of Engineers*, 1995, p. 211-215) fornece as informações a seguir sobre as distâncias (km) de um determinado sistema:

Extensão	6–<8	8–<10	10–<12	12–<14	14–<16
Freqüência	6	23	30	35	32
Extensão	16–<18	18–<20	20–<22	22–<24	24–<26
Freqüência	48	42	40	28	27
Extensão	26–<28	28–<30	30–<35	35–<40	40–<45
Freqüência	26	14	27	11	2

a. Desenhe um histograma correspondente a estas freqüências.

b. Que proporção destas distâncias são inferiores a 20? Que proporção das rotas possui distância ao menos de 30?

c. *Grosso modo*, qual é o valor do 90º percentil da distribuição de distâncias de rotas?

d. *Grosso modo*, qual é a mediana da distância de rota?

81. Um estudo realizado para investigar a distribuição do tempo total de frenagem (tempo de reação mais tempo do movimento entre acelerador e freio, em ms) em condições reais de dirigibilidade a 60km/h forneceu as informações resumidas sobre a distribuição dos tempos ("A Field Study on Braking Responses during Driving," *Ergonomics*, 1995: 1903-1910):

média = 535 mediana = 500 moda = 500
dp = 96 mínimo = 220 máximo = 925
5º percentil = 400 10º percentil = 430
90º percentil = 640 95º percentil = 720

O que você pode concluir sobre o formato do histograma desses dados? Explique seu raciocínio.

82. Os dados de amostra $x_1, x_2, \ldots, x_n$ algumas vezes representam uma **série temporal**, onde x_t = valor observado de uma variável de resposta x no momento t. Freqüentemente, a série observada mostra grande variação aleatória, o que dificulta o estudo do comportamento de longo prazo. Nessas situações, é desejável produzir uma versão suavizada da série. Uma das técnicas usadas é a suavização **exponencial**. O valor de uma constante de suavização α é escolhido ($0 < \alpha < 1$). Então, com $\bar{x}_t$ = valor ajustado no instante t, definimos $\bar{x}_t = x_1$ e para $t = 2, 3, \ldots, n$, $\bar{x}_t = \alpha x_t + (1-\alpha)\bar{x}_{t-1}$.

a. Considere a seguinte série de tempo em que x_t = temperatura (°F) de efluente em uma estação de tratamento de esgoto no dia t: 47, 54, 53, 50, 46, 46, 47, 50, 51, 50, 46, 52, 50, 50. Plote cada coordenada x_t em relação a t em um sistema bidimensional (um gráfico de série de tempo). Parece haver algum padrão?

b. Calcule o $\bar{x}_t$ usando $\alpha = 0{,}1$. Repita, para $\alpha = 0{,}5$. Que valor de α fornece uma série $\bar{x}_t$ mais suavizada?

c. Substitua $\bar{x}_{t-1} = \alpha x_{t-1} + (1-\alpha)\bar{x}_{t-2}$ no lado direito da expressão de $\bar{x}_t$ e então substitua $\bar{x}_{t-2}$ em termos de x_{t-2}, $\bar{x}_{t-3}$ e assim por diante. De quantos valores $x_t, x_{t-1}, \ldots, x_1$, $\bar{x}_t$ depende? O que acontece ao coeficiente de x_{t-k} quando k aumenta?

d. Consulte a parte (c). Se t for grande, qual é a sensibilidade de $\bar{x}_t$ em relação ao valor inicial $\bar{x}_t = x_1$? Explique.

(*Nota:* Uma referência relevante é o artigo "Simple Statistics for Interpreting Environmental Data," *Water Pollution Control Fed. J.*, 1981, p. 167-175.)

83. Considere as observações numéricas $x_1, \ldots, x_n$. Freqüentemente é de interesse saber se os x_i estão (ao menos aproximadamente) distribuídos simetricamente ao redor de algum valor. Se n é pelo menos moderadamente grande, a extensão da simetria pode ser avaliada por meio de um diagrama de caule e folha ou de um histograma. Entretanto, se n não for muito grande, esses gráficos não serão muito informativos. Considere a alternativa a seguir: seja y_1 o menor x_i, y_2 o segundo menor x_i e assim por diante. Faça então o gráfico dos seguintes pares como pontos em um sistema de coordenadas bidimensional: $(y_n - \tilde{x}, \tilde{x} - y_1)$, $(y_{n-1} - \tilde{x}, \tilde{x} - y_2)$, $(y_{n-2} - \tilde{x}, \tilde{x} - y_3)$, ... Há $n/2$ pontos em que n é par e $(n-1)/2$ em que n é ímpar.

a. Qual é a aparência do gráfico quando há simetria perfeita dos dados? Qual sua aparência quando as observações se estendem mais acima do que abaixo da mediana (uma cauda superior longa)?

b. Os dados a seguir sobre precipitação (acre-pé) de 26 nuvens esparsas foram obtidos do artigo "A Bayesian Analysis of a Multiplicative Treatment Effect in Weather Modification" (*Technometrics*, 1975, p. 161-166). Construa o gráfico e comente sobre a extensão da simetria ou a natureza do seu desvio.

4,1	7,7	17,5	31,4	32,7	40,6	92,4
115,3	118,3	119,0	129,6	198,6	200,7	242,5
255,0	274,7	274,7	302,8	334,1	430,0	489,1
703,4	978,0	1656,0	1697,8	2745,6		

Bibliografia

CHAMBERS, John. CLEVELAND, William, KLEINER, Beat e TUKEY, Paul. *Graphical Methods for Data Analysis*. Brooks/Cole, Pacific Grove, CA, 1983. Uma apresentação altamente recomendada sobre metodologia gráfica antiga e mais recente da estatística.

DEVORE, Jay e PECK, Roxy. *Statistics: The Exploration and Analysis of Data* (4ª ed.). Duxbury Press, Pacific Grove, CA, 2001. Os primeiros capítulos fornecem uma pesquisa não-matemática de métodos para descrição e resumo de dados.

FREEDMAN, David, PISANI, Robert e PURVES, Roger. *Statistics* (3ª ed.). Norton, Nova York, 1998. Uma excelente pesquisa não-matemática de raciocínio e metodologia básicos da estatística.

HOAGLIN, David, MOSTELLER, Frederick e TUKEY, John. *Understanding Robust and Exploratory Data Analysis*. Wiley, Nova York, 1983. Discute por que e como os métodos exploratórios devem ser empregados. É bastante útil em detalhes de diagramas de caule e folha e *boxplots*.

HOAGLIN, David e VELLEMAN, Paul. *Applications, Basics, and Computing of Exploratory Data Analysis*. Duxbury Press, Boston, 1980. Uma boa discussão de alguns métodos exploratórios básicos.

MOORE, David. *Statistics: Concepts and Controversies* (5ª ed.). Freeman, San Francisco, 2001. Um estudo extremamente agradável e de fácil leitura que contém uma discussão intuitiva de problemas relacionados à amostragem e experimentos projetados.

TANUR, Judith et al. (eds.). *Statistics: A Guide to the Unknown* (3ª ed.). Duxbury Press, Belmont, CA, 1988. Contém vários artigos não-técnicos sobre diversas aplicações da estatística.

2

Probabilidade

Introdução

O termo **probabilidade** se refere ao estudo da aleatoriedade e da incerteza.

Em qualquer situação em que ocorrem diversos resultados, a teoria da probabilidade oferece métodos de quantificação das chances ou possibilidades de ocorrência associadas aos diversos resultados. A linguagem da probabilidade é constantemente usada de maneira informal em contextos escritos e falados. Exemplos incluem declarações como: "é provável que a média Dow-Jones cresça no final do ano", "há 50% de chance de o titular buscar a reeleição", "provavelmente pelo menos uma seção deste curso será oferecida no ano que vem", "as chances favorecem um acordo rápido para o fim da greve" e "espera-se que pelo menos 20 mil ingressos sejam vendidos para o show". Neste capítulo, apresentaremos alguns conceitos básicos de probabilidade, indicaremos como as probabilidades podem ser interpretadas e como suas regras podem ser aplicadas para calcular as possibilidades de ocorrência de muitos eventos. A metodologia da probabilidade permitirá que expressemos, em linguagem precisa, declarações informais como as fornecidas acima.

O estudo da probabilidade como ramo da matemática data de mais de 300 anos e teve sua gênese relacionada a questões que envolviam jogos de azar. Muitos livros se dedicam exclusivamente à probabilidade, mas nosso objetivo é cobrir apenas a parte da disciplina que tem maior ligação com os problemas de inferência estatística.

2.1 Espaços amostrais e eventos

Um **experimento** é qualquer ação ou processo cujo resultado está sujeito à incerteza. Apesar de o termo *experimento* geralmente sugerir uma situação altamente controlada de teste em laboratório, aqui será usado em um sentido muito mais amplo. Dessa forma, os experimentos que são de interesse incluem jogar uma moeda uma ou diversas vezes, selecionar uma ou várias cartas de um baralho, pesar um pedaço de pão, determinar o tempo de

viagem de casa ao trabalho em uma dada manhã, obter tipos sangüíneos de um grupo de indivíduos ou medir as resistências de compressão de diversas vigas metálicas.

O espaço amostral de um experimento

DEFINIÇÃO

O **espaço amostral** de um experimento, representado por S, é o conjunto de todos os resultados possíveis desse experimento.

Exemplo 2.1

O experimento mais simples a que a probabilidade se aplica é aquele com dois resultados finais possíveis. Um exemplo de experimento desse tipo é examinar um único fusível para ver se funciona. O espaço amostral desse experimento é expresso como $\mathcal{S} = \{N, D\}$, onde N representa sem defeito, D representa com defeito e as chaves são usadas para englobar os elementos do conjunto. Outro experimento desse tipo envolve jogar um percevejo e observar se sua ponta cai para cima ou para baixo, com o espaço amostral $\mathcal{S} = \{C, B\}$ e um terceiro consiste em observar o sexo da próxima criança nascida no hospital local, com $\mathcal{S} = \{M, F\}$.

Exemplo 2.2

Se examinarmos três fusíveis em seqüência e anotarmos o resultado de cada exame, o resultado do experimento é qualquer seqüência de N e D de 3 elementos, de forma que

$$\mathcal{S} = \{NNN, NND, NDN, NDD, DNN, DND, DDN, DDD\}$$

Se tivéssemos jogado um percevejo três vezes, o espaço amostral seria obtido substituindo N por C no $\mathcal{S}$ acima, e com uma alteração similar na notação do espaço amostral para o experimento em que o sexo de três crianças recém-nascidas é observado. ■

Exemplo 2.3

Dois postos de gasolina estão localizados em uma determinada interseção. Cada um possui seis bombas. Considere o experimento em que o número de bombas em uso em determinada hora do dia é determinado para cada posto. Um resultado experimental especifica quantas bombas estão em uso no primeiro posto e quantas são usadas no segundo. Um resultado possível é (2, 2), outro é (4, 1) e outro ainda é (1, 4). Os 49 resultados em $\mathcal{S}$ são exibidos na tabela a seguir. O espaço amostral do experimento em que um dado de seis lados é lançado duas vezes é obtido excluindo a linha 0 e a coluna 0 da tabela, fornecendo 36 resultados.

					Segundo Posto			
		0	**1**	**2**	**3**	**4**	**5**	**6**
	0	(0, 0)	(0, 1)	(0, 2)	(0, 3)	(0, 4)	(0, 5)	(0, 6)
	1	(1, 0)	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
	2	(2, 0)	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
Primeiro posto	**3**	(3, 0)	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
	4	(4, 0)	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
	5	(5, 0)	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
	6	(6, 0)	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

■

Exemplo 2.4

Se uma bateria de lanterna nova, tipo D, tiver uma voltagem fora de certos limites, será classificada como falha (*F*); se a voltagem estiver dentro dos limites especificados, será classificada como sucesso (*S*). Suponha que um experimento consista em testar cada bateria quando sai de uma linha de montagem até que seja observado um sucesso. Apesar de não ser muito provável, um resultado possível é que as primeiras 10 (ou 100 ou 1000 ou ...) sejam *F* e a próxima seja *S*. Isto é, para qualquer inteiro *n* possível, teremos de examinar *n* baterias antes de obter o primeiro *S*. O espaço amostral é $\mathcal{S} = \{S, FS, FFS, FFFS, \ldots\}$, que contém um número infinito de resultados possíveis. A mesma forma abreviada de espaço amostral é apropriada para um experimento em que, é registrado o sexo de cada criança recém-nascida, em um horário especificado, até que seja observado o nascimento de uma criança do sexo masculino. ■

Eventos

No estudo de probabilidade, estaremos interessados não apenas nos resultados individuais de $\mathcal{S}$, como também em qualquer grupo de resultados de $\mathcal{S}$.

DEFINIÇÃO **Evento** é qualquer grupo (subconjunto) de resultados contidos no espaço amostral $\mathcal{S}$. O evento é denominando **simples** se consistir um único resultado e **composto** se consistir em mais de um resultado.

Quando um experimento é realizado, determinado evento *A* ocorre se o resultado experimental estiver contido em *A*. Em geral, ocorrerá exatamente um evento simples, mas diversos eventos compostos também podem ocorrer simultaneamente.

Exemplo 2.5

Considere um experimento em que cada um de três veículos que trafeguem em uma determinada estrada siga pela saída à esquerda *(E)* ou à direita *(D)* no final da rampa de saída. Os oito resultados possíveis que compõem o espaço amostral são: *EEE, DEE, EDE, EED, EDD, DED, DDE e DDD*. Dessa forma, há oito eventos simples, dentre os quais estão $E_1 = \{EEE\}$ e $E_5 = \{EDD\}$. Os eventos compostos incluem:

$A = \{DEE, EDE, EED\}$ = o evento em que exatamente um dos três veículos vira à direita;

$B = \{EEE, DEE, EDE, EED\}$ = o evento em que no máximo um dos veículos vira à direita;

$C = \{EEE, DDD\}$ = o evento em que os três veículos viram na mesma direção.

Suponha que, quando o experimento é executado, o resultado seja *EEE*. Então, o evento simples E_1 terá ocorrido, da mesma forma que os eventos *B* e *C* (mas não *A*). ■

Exemplo 2.6 (continuação do Exemplo 2.3)

Quando o número de bombas em uso em cada um dos dois postos de seis bombas for observado, haverá 49 resultados possíveis, de forma que haverá 49 eventos simples: $E_1 = \{(0, 0)\}$, $E_2 = \{(0, 1)\}, \ldots, E_{49} = \{(6, 6)\}$. Exemplos de eventos compostos são

$A = \{(0, 0), (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$ = o evento em que o número de bombas em uso é o mesmo nos dois postos;

$B = \{(0, 4), (1, 3), (2, 2), (3, 1), (4, 0)\}$ = o evento em que o número total de bombas em uso é quatro;

$C = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ = o evento em que no máximo uma bomba está em uso em cada posto. ■

Exemplo 2.7 (continuação do Exemplo 2.4)

O espaço amostral do experimento de exame das baterias contém um número infinito de resultados, portanto há um número infinito de eventos simples. Os eventos compostos incluem:

$A = \{S, FS, FFS\}$ = o evento em que no máximo três baterias são examinadas;

$E = \{FS, FFFS, FFFFFS, ...\}$ = o evento em que um número par de baterias é examinado. ■

Algumas relações sobre a teoria dos conjuntos

Um evento é essencialmente um conjunto, de forma que as relações e resultados da teoria elementar dos conjuntos podem ser usados para o estudo dos eventos. As operações a seguir serão usadas para construção de novos eventos, a partir de eventos conhecidos.

DEFINIÇÃO

1. A **união** de dois eventos A e B, representada por $A \cup B$ e lida *"A união B"*, é o evento que consiste em todos os resultados que estão no evento A *ou no B ou em ambos* (de forma que a união inclui resultados em que ocorram A e B, bem como aqueles em que exatamente um ocorre), isto é, todos os resultados em ao menos um dos eventos.
2. A **interseção** dos dois eventos A e B, representada por $A \cap B$ e lida *"A interseção B"*, é o evento que consiste de todos os resultados que estão em *ambos A e B*.
3. O **complemento** de um evento A, representado por A', é o conjunto de todos os resultados em $\mathcal{S}$ que não estão contidos em A.

Exemplo 2.8 (continuação do Exemplo 2.3)

Para o experimento em que é observado o número de bombas em uso em um posto de gasolina de seis bombas, assuma $A = \{0, 1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$ e $C = \{1, 3, 5\}$. Então

$A \cup B = \{0, 1, 2, 3, 4, 5, 6\} = \mathcal{S}$; $A \cup C = \{0, 1, 2, 3, 4, 5\}$,

$A \cap B = \{3, 4\}$, $A \cap C = \{1, 3\}$, $A' = \{5, 6\}$, $\{A \cup C\}' = \{6\}$ ■

Exemplo 2.9 (continuação do Exemplo 2.4)

No experimento da bateria, defina A, B e C por

$A = \{S, FS, FFS\}$

$B = \{S, FFS, FFFFS\}$

e

$C = \{FS, FFFS, FFFFFS, ...\}$

Então

$A \cup B = \{S, FS, FFS, FFFFS\}$

$A \cap B = \{S, FFS\}$

$A' = \{ FFFS, FFFFS, FFFFFS, ...\}$

e

$C' = \{S, FFS, FFFFS, ...\}$ = {evento em que é inspecionado um número ímpar de baterias} ■

Algumas vezes, A e B não possuem resultados em comum, de modo que a interseção de A e B não contém nenhum resultado.

DEFINIÇÃO Quando A e B não possuem resultados em comum, são chamados eventos **mutuamente exclusivos** ou **disjuntos**.

Exemplo 2.10

Uma cidade pequena possui três revendedores de automóveis: um revendedor GM que vende Chevrolets, Pontiacs e Buicks, um revendedor Ford que vende Fords e Mercurys e um revendedor Chrysler que vende Plymouths e Chryslers. Se um experimento consistir em observar a marca do próximo carro vendido, os eventos $A = \{$Chevrolet, Pontiac, Buick$\}$ e $B = \{$Ford, Mercury$\}$ são mutuamente exclusivos porque o próximo carro vendido não pode ser um produto GM e Ford ao mesmo tempo. ■

As operações de união e interseção podem ser estendidas a mais de dois eventos. Para quaisquer três eventos A, B e C, o evento $A \cup B \cup C$ é o conjunto de resultados contidos em no mínimo um dos três eventos, enquanto $A \cap B \cap C$ é o conjunto de resultados contidos nos três eventos. Dados os eventos $A_1, A_2, A_3, \ldots$, eles serão denominados mutuamente exclusivos (ou disjuntos em pares) se dois deles não tiverem resultados em comum.

Uma representação gráfica de eventos e manipulações de eventos é obtida pelo uso de diagramas de Venn. Para construir um desses diagramas, desenhe um retângulo cujo interior representará o espaço amostral $\mathcal{S}$. Então, qualquer evento A é representado como o interior de uma curva fechada (normalmente um círculo) contido em $\mathcal{S}$. A Figura 2.1 mostra exemplos de diagramas de Venn.

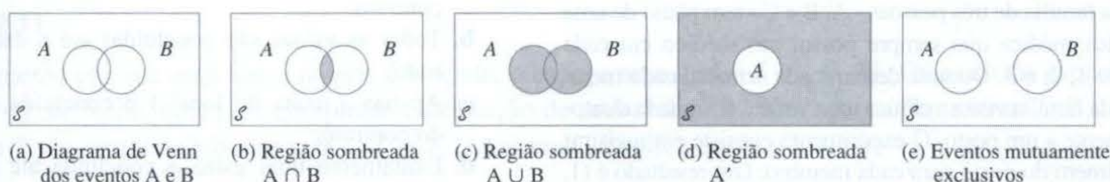


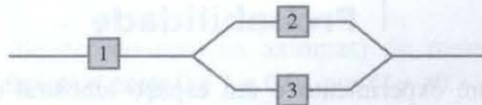
Figura 2.1 Diagramas de Venn

Exercícios Seção 2.1 (1–10)

1. Quatro universidades — 1, 2, 3 e 4 — estão participando de um torneio de basquete. Na primeira *etapa*, 1 jogará com 2 e 3 com 4. Os dois vencedores disputarão o campeonato e os dois perdedores também jogarão. Um resultado possível pode ser representado por 1324 (1 ganha de 2 e 3 ganha de 4 nos jogos da primeira *etapa* e 1 ganha de 3 e 2 ganha de 4).
   a. Relacione todos os resultados de $\mathcal{S}$.
   b. Represente por A o evento em que 1 ganha o torneio. Relacione os resultados de A.
   c. Represente por B o evento em que 2 seja um dos finalistas do campeonato. Relacione os resultados de B.
   d. Quais são os resultados de $A \cup B$ e de $A \cap B$? Quais são os resultados de A'?

2. Suponha que os veículos que trafegam em uma determinada estrada possam tomar uma saída à direita (D), à esquerda (E) ou ir em frente (F). Observe a direção de cada um de três veículos sucessivamente.
   a. Relacione todos os resultados do evento A em que os três veículos seguem na mesma direção.
   b. Relacione todos os resultados do evento B em que os três veículos tomam diferentes direções.
   c. Relacione os resultados do evento C em que exatamente dois dos três veículos viram à direita.
   d. Relacione todos os resultados do evento D em que exatamente dois veículos seguem na mesma direção.
   e. Relacione os resultados em D', $C \cup D$, e $C \cap D$.

3. Três componentes estão conectados para formar um sistema conforme exibido no diagrama a seguir. Como os componentes no subsistema 2-3 estão conectados em paralelo, esse subsistema funcionará se ao menos um dos dois componentes individuais funcionar. Para que todo o sistema funcione, o componente 1 deve funcionar, bem como o subsistema 2-3.



O experimento consiste em determinar a condição de cada componente [S (sucesso) para um componente que funciona bem e F (falha) para um componente que não funciona].

a. Que resultados estão contidos no evento A para que exatamente dois dos três componentes funcionem?

b. Que resultados estão contidos no evento B para que ao menos dois componentes funcionem?

c. Que resultados estão contidos no evento C para que o sistema funcione?

d. Relacione os resultados de C', $A \cup C$, $A \cap C$, $B \cup C$, e $B \cap C$.

4. Cada item de uma amostra de quatro hipotecas está classificado como taxa fixa (F) ou taxa variável (V).

a. Quais são os 16 resultados de $\mathcal{S}$?

b. Que resultados do evento possuem exatamente três hipotecas de taxa fixa?

c. Que resultados do evento possuem quatro hipotecas de taxa fixa?

d. Que resultados pertencem ao evento em que as quatro hipotecas são de taxa variável?

e. Qual é a união dos eventos das partes (c) e (d) e qual a interseção desses dois eventos?

f. Quais são a união e a interseção dos dois eventos das partes (b) e (c)?

5. Uma família de três pessoas –A, B e C– tem plano de uma clínica médica que sempre possui um médico em cada posto 1, 2, e 3. Durante determinada semana, cada membro da família visita a clínica uma vez e é designada aleatoriamente a um posto. O experimento consiste em registrar o número do posto para cada membro. Um resultado é (1, 2, 1) para A no posto 1, B no posto 2 e C no posto 1.

a. Relacione os 27 resultados do espaço amostral.

b. Relacione todos os resultados do evento em que os três membros seguem para o mesmo posto.

c. Relacione todos os resultados do evento em que todos os membros seguem para postos diferentes.

d. Relacione todos os resultados do evento em que ninguém segue para o posto 2.

6. A biblioteca de uma faculdade possui cinco exemplares de certo livro-texto de reserva. Duas cópias (1 e 2) são primeiras edições e as outras três (3, 4 e 5) são segundas edições. Um aluno examina esses livros em ordem aleatória, parando apenas quando uma segunda edição é selecionada. Um resultado possível é 5 e outro é 213.

a. Relacione todos os resultados de $\mathcal{S}$.

b. Represente por A o evento em que exatamente um livro deve ser examinado. Quais são os resultados em A?

c. Represente por B o evento em que o livro 5 é o selecionado. Quais resultados pertencem a B?

d. Represente por C o evento em que o livro 1 não é examinado. Quais resultados pertencem a C?

7. Um departamento acadêmico terminou a votação secreta para sua chefia. A urna contém quatro cédulas com votos para o candidato A e três com votos para o candidato B. Suponha que essas cédulas sejam removidas da urna uma a uma.

a. Relacione todos os resultados possíveis.

b. Suponha que uma contagem seja feita conforme as cédulas são removidas. Para que resultados A se mantém na frente de B na contagem?

8. Uma empresa de engenharia civil está trabalhando atualmente em usinas de energia em três locais diferentes. Represente por A_i o evento em que a usina no local i é completada na data do contrato. Use as operações de união, interseção e complemento para descrever cada um dos eventos a seguir nos termos de A_1, A_2 e A_3, desenhe um diagrama de Venn e sombreie a região correspondente a cada uma.

a. Ao menos uma usina é concluída até a data do contrato.

b. Todas as usinas são concluídas até a data do contrato.

c. Apenas a usina do local 1 é concluída até a data do contrato.

d. Exatamente uma usina é concluída até a data do contrato.

e. Apenas a usina do local 1 ou as duas outras são concluídas até a data do contrato.

9. Use diagramas de Venn para provar as duas relações a seguir para eventos A e B (denominadas leis de De Morgan) para quaisquer eventos A e B:

a. $(A \cup B)' = A' \cap B'$

b. $(A \cap B)' = A' \cup B'$

10. a. No Exemplo 2.10, identifique três eventos que não são mutuamente exclusivos.

b. Suponha que não haja um resultado comum a todos os três eventos A, B e C. Esses eventos são necessariamente mutuamente exclusivos? Se sua resposta for sim, explique. Se for não, dê um exemplo contrário, usando o experimento do Exemplo 2.10.

2.2 Axiomas, Interpretações e Propriedades da Probabilidade

Dados um experimento e um espaço amostral $\mathcal{S}$, *o objetivo da probabilidade é atribuir a cada evento A um número P(A), denominado probabilidade do evento A, que fornecerá uma medida precisa da chance de*

ocorrência de A. Para assegurar que as atribuições de probabilidade sejam consistentes com nossas noções intuitivas de probabilidade, todas as atribuições devem satisfazer os axiomas a seguir (propriedades básicas) de probabilidade.

AXIOMA 1 Para qualquer evento, A, $P(A) \geq 0$.

AXIOMA 2 $P(\mathcal{S}) = 1$.

AXIOMA 3 **a.** Se $A_1, A_2, \ldots, A_k$ *for um* conjunto finito de eventos mutuamente exclusivos, então

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) = \sum_{i=1}^{k} P(A_i)$$

b. Se $A_1, A_2, A_3, \ldots$ for um conjunto infinito de eventos mutuamente exclusivos, então

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i)$$

O Axioma 1 reflete a noção intuitiva de que a chance de ocorrência de A deve ser não-negativa. O espaço amostral é, por definição, o evento que deve ocorrer quando o experimento é realizado ($\mathcal{S}$ contém todos os resultados possíveis), de forma que o Axioma 2 diz que a maior probabilidade possível de ser atribuída a $\mathcal{S}$ é 1. O terceiro axioma formaliza a idéia de que, se desejarmos que a probabilidade de ao menos um de diversos eventos ocorra e que dois eventos não ocorram simultaneamente, a chance de pelo menos um ocorrer será a soma das chances dos eventos individuais.

Exemplo 2.11

No experimento em que uma única moeda é lançada, o espaço amostral é $\mathcal{S} = \{H, T\}$. Os axiomas especificam que $P(\mathcal{S}) = 1$, de forma que, para completar a atribuição de probabilidade, falta apenas determinar $P(H)$ e $P(T)$. Já que H e T são eventos disjuntos e $H \cup T = \mathcal{S}$, o Axioma 3 implica que

$$1 = P(\mathcal{S}) = P(H) + P(T)$$

Essa expressão implica que $P(T) = 1 - P(H)$. A única liberdade permitida pelos axiomas nesses experimentos é a probabilidade atribuída a H. Uma possível atribuição de probabilidades é $P(H) = 0{,}5$, $P(T) = 0{,}5$, enquanto outra atribuição possível é $P(H) = 0{,}75$, $P(T) = 0{,}25$. De fato, representar p por qualquer número fixo entre 0 e 1, $P(H) = p$ e $P(T) = 1 - p$ é uma atribuição consistente com os axiomas. ■

Exemplo 2.12

Considere o experimento do Exemplo 2.4, em que as baterias que saem de uma linha de montagem são testadas uma a uma até que seja encontrada uma com voltagem dentro dos limites especificados. Esses eventos simples são $E_1 = \{S\}$, $E_2 = \{FS\}$, $E_3 = \{FFS\}$, $E_4 = \{FFFS\}$, ... Suponha que a probabilidade de uma bateria ser satisfatória seja de 0,99. Então, pode ser demonstrado que $P(E_1) = 0{,}99$, $P(E_2) = (0{,}01)(0{,}99)$, $P(E_3) = (0{,}01)^2(0{,}99)$... é uma atribuição de probabilidades a eventos simples que satisfazem os axiomas. Em particular, porque os E_i são disjuntos e $\mathcal{S} = E_1 \cup E_2 \cup E_3 \cup \ldots$, deve ser o caso de que

$$1 = P(\mathcal{S}) = P(E_1) + P(E_2) + P(E_3) + \cdots$$
$$= 0{,}99[1 + 0{,}01 + (0{,}01)^2 + (0{,}01)^3 + \cdots]$$

A validade dessa igualdade é uma conseqüência de um resultado matemático referente à soma de uma série geométrica.

Entretanto, outra atribuição de probabilidade legítima (de acordo com os axiomas) do mesmo tipo "geométrico" é obtida pela substituição de 0,99 por qualquer número p entre 0 e 1 e 0,01 por $(1 - p)$. ■

Interpretando probabilidade

Os Exemplos 2.11 e 2.12 mostram que os axiomas não determinam completamente uma atribuição de probabilidades a eventos. Os axiomas servem apenas para cortar atribuições inconsistentes com as noções intuitivas de probabilidade. No experimento de lançamento de uma moeda do Exemplo 2.11, duas atribuições particulares foram sugeridas. A atribuição correta ou apropriada depende da forma pela qual o experimento é executado e também da interpretação da probabilidade. A interpretação usada com mais freqüência e mais compreendida baseia-se na noção de freqüências relativas.

Considere um experimento que pode ser executado repetidamente de uma forma idêntica e independente e represente por *A* um evento que consista de um conjunto fixo de resultados do experimento. Exemplos simples de experimentos passíveis de repetição incluem o lançamento de dados e moedas discutidos anteriormente. Se o experimento for executado *n* vezes, em algumas delas o evento *A* ocorrerá (o resultado será o conjunto *A*) e em outras, não. Represente por $n(A)$ o número de repetições em que *A* ocorre. A relação $n(A)/n$ é denominada *freqüência relativa* de ocorrência do evento *A* na seqüência de *n* repetições. Evidências empíricas, com base nos resultados de várias dessas seqüências de experimentos repetidos, indicam que, conforme *n* aumenta, a freqüência relativa $n(A)/n$ se estabiliza, de acordo com ilustração da Figura 2.2. Isto é, conforme *n* torna-se arbitrariamente grande, a freqüência reativa se aproxima de um valor-limite denominado *freqüência relativa-limite* do evento *A*. A interpretação objetiva da probabilidade identifica essa freqüência relativa-limite com $P(A)$.

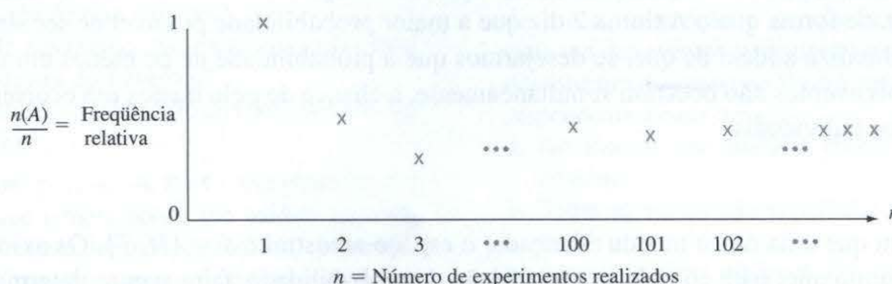


Figura 2.2 Estabilização da freqüência relativa

Se forem atribuídas probabilidades a eventos de acordo com suas freqüências relativas-limite, podemos interpretar a afirmativa "a probabilidade de dar 'cara' em um lançamento de moeda é de 0,5" para indicar que, em grande número de lançamentos, a "cara" será exibida em cerca de metade deles e "coroa", na outra metade.

A interpretação de freqüência relativa da probabilidade é tida como objetiva porque toma como base uma propriedade do experimento, em vez de qualquer indivíduo preocupado com o experimento. Por exemplo: dois observadores de uma seqüência de lançamentos de moeda devem usar as mesmas atribuições de probabilidade, já que eles não influenciam a freqüência relativa-limite. Na prática, essa interpretação não é tão objetiva quanto parece, já que a freqüência relativa-limite de um evento não será conhecida. Portanto, temos de atribuir probabilidades com base em nossa crença sobre a freqüência relativa-limite dos eventos em estudo. Felizmente, há muitos experimentos para os quais há consenso no que diz respeito a atribuições de probabilidade. Quando falamos de uma moeda justa, queremos dizer que $P(H) = P(T) = 0{,}5$; e um dado justo é aquele em que as freqüências relativas-limite dos seis resultados são todas $\frac{1}{6}$, sugerindo atribuições de probabilidade $P(\{1\}) = \cdots = P(\{6\}) = \frac{1}{6}$.

Como a interpretação objetiva da probabilidade se baseia na noção de freqüência-limite, sua aplicabilidade é limitada a situações experimentais que podem ser repetidas. Ainda assim, a linguagem de probabilidade é normalmente usada em relação a situações que não podem ser repetidas. Os exemplos incluem: "há boas chances de um acordo de paz", "provavelmente nossa empresa fechará o contrato" e "como o melhor atacante deles está machucado, espero que eles não marquem mais de 10 contra nós". Nessas situações, desejamos, como antes, atribuir probabilidades numéricas a diversos resultados e eventos (por exemplo: a probabilidade de fecharmos o contrato é de 0,9). Devemos, portanto, adotar uma interpretação alternativa dessas probabilidades. Como

diferentes observadores podem ter diversas informações prévias e opiniões a respeito de situações experimentais, as atribuições de probabilidade podem ser diferentes de indivíduo para indivíduo. As interpretações nestas situações são denominadas *subjetivas*. O livro de Robert Winkler relacionado nas referências do capítulo fornece uma boa pesquisa sobre diversas interpretações subjetivas.

Propriedades de probabilidade

PROPOSIÇÃO Para qualquer evento, A, $P(A) = 1 - P(A')$.

Demonstração

No Axioma 3a, assuma $k = 2$, $A_1 = A$, e $A_2 = A'$. Já que por definição A', $A \cup A' = \mathcal{S}$ enquanto A e A' são disjuntos, $1 = P(\mathcal{S}) = P(A \cup A') = P(A) + P(A')$, de onde segue o resultado. ■

Essa proposição é surpreendentemente útil, porque há várias situações em que $P(A')$ é mais facilmente obtido pelos métodos diretos do que $P(A)$.

Exemplo 2.13

Considere um sistema de cinco componentes idênticos ligados em série, conforme ilustrado na Figura 2.3.

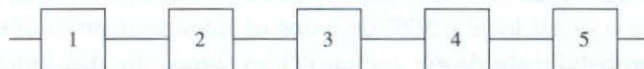


Figura 2.3 Um sistema de cinco componentes ligados em série

Represente um componente que falha por F e o que não falha por S (de sucesso). Represente por A o evento em que o *sistema* falha. Para que A ocorra, ao menos um dos componentes individuais deve falhar. Os resultados de A incluem $SSFSS$ (1, 2, 4 e 5 funcionam, mas 3 não), $FFSSS$ e assim por diante. Na verdade há 31 diferentes resultados em A. Entretanto, A', o evento em que o sistema funciona, consiste em um único resultado $SSSSS$. Na Seção 2.5 veremos que se 90% de todos esses componentes não apresentarem falhas e se componentes diferentes apresentarem falhas independentemente um do outro, $P(A') = P(SSSSS) = 0{,}9^5 = 0{,}59$. Portanto, $P(A) = 1 - 0{,}59 = 0{,}41$. Dessa forma, em um grande número de tais sistemas, cerca de 41% apresentarão falhas. ■

Em geral, a proposição acima é útil quando o evento de interesse pode ser expresso como "ao menos...", já que o complemento "menos que..." pode ser mais fácil de se lidar (em alguns problemas, "mais que..." é mais fácil de lidar que "no máximo..."). Se você tiver dificuldade para calcular $P(A)$ diretamente, pense em determinar $P(A')$.

PROPOSIÇÃO Se A e B forem mutuamente exclusivos, então $P(A \cap B) = 0$.

Demonstração

Visto que $A \cap B$ não contém resultados, $(A \cap B)' = \mathcal{S}$. Portanto, $1 = P[(A \cap B)'] = 1 - P(A \cap B)$, o que implica $P(A \cap B) = 1 - 1 = 0$. ■

Quando os eventos A e B são mutuamente exclusivos, o Axioma 3 fornece $P(A \cup B) = P(A) + P(B)$. Quando A e B não são mutuamente exclusivos, a probabilidade da união é obtida do resultado a seguir.

PROPOSIÇÃO Para quaisquer dois eventos A e B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Observe que a proposição é válida mesmo se A e B forem mutuamente exclusivos, já que $P(A \cap B) = 0$. A idéia central é que, ao adicionar $P(A)$ e $P(B)$, a probabilidade da interseção $A \cap B$ será contada duas vezes, de forma que $P(A \cap B)$ deve ser subtraído.

Demonstração

Observe primeiro que $A \cup B = A \cup (B \cap A')$, conforme ilustrado na Figura 2.4. Visto que A e $(B \cap A')$ são mutuamente exclusivos, $P(A \cup B) = P(A) + P(B \cap A')$. Mas $B = (B \cap A) \cup (B \cap A')$ (a união da parte de B contida em A e a parte de B não contida em A), sendo $(B \cap A)$ e $(B \cap A')$ mutuamente exclusivos, de forma que $P(B) = P(B \cap A) + P(B \cap A')$. Combinando esses resultados, temos

$$P(A \cup B) = P(A) + P(B \cap A') = P(A) + [P(B) - P(A \cap B)]$$
$$= P(A) + P(B) - P(A \cap B)$$

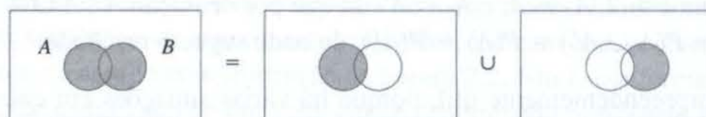


Figura 2.4 Representação de $A \cup B$ como união de eventos disjuntos ■

Exemplo 2.14

Em um determinado bairro residencial, 60% de todos os lares assinam o jornal metropolitano publicado em uma cidade próxima, 80% assinam o jornal local e 50% de todos os lares assinam os dois. Se um lar for selecionado aleatoriamente, qual será a probabilidade de ele assinar (1) ao menos um dos jornais e (2) exatamente um dos dois jornais?

Fazendo $A = \{$assinaturas do jornal metropolitano$\}$ e $B = \{$assinaturas do jornal local$\}$, as informações fornecidas implicam $P(A) = 0{,}6$, $P(B) = 0{,}8$, e $P(A \cap B) = 0{,}5$. A proposição anterior se aplica a dar

P(assinaturas de ao menos um dos jornais)

$$= P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0{,}6 + 0{,}8 - 0{,}5 = 0{,}9$$

O evento de um lar assinar apenas o jornal local pode ser escrito como $A' \cap B$ [(não-metropolitano) e local]. A Figura 2.4 implica que

$$0{,}9 = P(A \cup B) = P(A) + P(A' \cap B) = 0{,}6 + P(A' \cap B)$$

pela qual $P(A' \cap B) = 0{,}3$. De forma similar, $P(A \cap B') = P(A \cup B) - P(B) = 0{,}1$. Como ilustrado na Figura 2.5, onde se vê que

$$P(\text{exatamente um}) = P(A \cap B') + P(A' \cap B) = 0{,}1 + 0{,}3 = 0{,}4$$

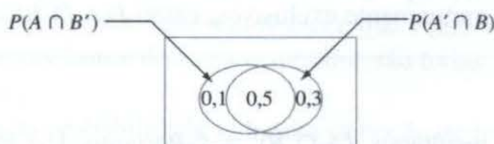


Figura 2.5 Probabilidades do Exemplo 2.14 ■

A probabilidade de uma união de mais de dois eventos pode ser calculada de forma análoga. Para os três eventos A, B e C, o resultado é

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Como vemos no diagrama de Venn de $A \cup B \cup C$, exibido na Figura 2.6, quando $P(A)$, $P(B)$ e $P(C)$ são adicionados, certas interseções são contadas duas vezes, de forma que devem ser subtraídas, mas isso resulta em $P(A \cap B \cap C)$ sendo subtraída mais de uma vez.

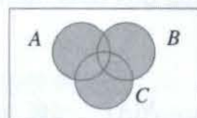


Figura 2.6 $A \cup B \cup C$

Determinando as sistemáticas de probabilidades

Quando o número de resultados possíveis (eventos simples) é grande, haverá muitos eventos compostos. Uma forma simples de determinar as probabilidades desses eventos, que evita a violação dos axiomas e as propriedades derivadas, é primeiro determinar as probabilidades $P(E_i)$ para todos os eventos simples. Isso deve satisfazer $P(E_i) \geq 0$ e $\sum_{\text{todos } i} P(E_i) = 1$. Dessa forma, a probabilidade de qualquer evento composto A é calculada pela adição das $P(E_i)$ para todos os E_i em A:

$$P(A) = \sum_{\text{todos os } E_i \text{ em } A} P(E_i)$$

Exemplo 2.15

Represente os seis eventos básicos $\{1\}, \ldots, \{6\}$ associados ao lançamento de um dado de seis lados uma vez por $E_1, \ldots E_6$. Se o dado for construído de tal forma que qualquer um dos três resultados pares tenha o dobro de probabilidade de ocorrer em relação aos ímpares, uma atribuição apropriada de probabilidades a eventos elementares é $P(E_1) = P(E_3) = P(E_5) = \frac{1}{9}$, $P(E_2) = P(E_4) = P(E_6) = \frac{2}{9}$. Dessa forma, para o evento $A = \{\text{resultado par}\} = E_2 \cup E_4 \cup E_6$, $P(A) = P(E_2) + P(E_4) + P(E_6) = \frac{6}{9} = \frac{2}{3}$; para $B = \{\text{resultado} \leq 3\} = E_1 \cup E_2 \cup E_3$, $P(B) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}$. ■

Resultados igualmente prováveis

Em muitos experimentos que consistem em N resultados, é razoável atribuir probabilidades iguais a todos os N eventos simples. Tais eventos incluem exemplos óbvios, como lançamento de uma moeda; ou um dado não viciado uma ou duas vezes (ou qualquer número fixo de vezes); ou selecionar uma ou diversas cartas de um baralho de 52 cartas bem embaralhado. Com $p = P(E_i)$ para cada i,

$$1 = \sum_{i=1}^{N} P(E_i) = \sum_{i=1}^{N} p = p \cdot N \qquad \text{então } p = \frac{1}{N}$$

Isto é, se houver N resultados possíveis, a probabilidade atribuída a cada um será $1/N$. Consideremos um evento A, com $N(A)$ representando o número de resultados contidos em A. Então

$$P(A) = \sum_{E_i \text{ em } A} P(E_i) = \sum_{E_i \text{ em } A} \frac{1}{N} = \frac{N(A)}{N}$$

Depois de contarmos o número N de resultados no espaço amostral para calcular a probabilidade de qualquer evento, devemos contar os resultados nele contidos e obter a razão entre os dois números. Dessa forma, quando os resultados forem igualmente prováveis, o cálculo das probabilidades se resume na contagem.

Exemplo 2.16

Quando dois dados são lançados separadamente, há $N = 36$ resultados (excluir as primeiras linha e coluna da tabela no Exemplo 2.3). Se os dois dados forem justos, todos os 36 resultados serão igualmente prováveis, então $P(E_i) = \frac{1}{36}$. Dessa forma, o evento $A = \{\text{soma dos dois números} = 7\}$ consistirá em seis resultados (1, 6), (2, 5), (3, 4), (4, 3), (5, 2) e (6, 1). Assim,

$$P(A) = \frac{N(A)}{N} = \frac{6}{36} = \frac{1}{6}$$ ■

Exercícios Seção 2.2 (11–28)

11. Uma empresa de fundos mútuos oferece a seus clientes diversos fundos: um de mercado, três de títulos diferentes, (curto, médio e longo prazos), dois fundos de ações (moderado e de alto risco) e um misto. Dentre os usuários que possuem cotas em apenas um fundo, seguem as porcentagens de clientes nos diferentes fundos.

Mercado	20%	Ações de alto risco	18%
Título curto prazo	15%	Ação de risco moderado	25%
Título intermediário	10%	Misto	7%
Título longo prazo	5%		

Um cliente que possui cotas em apenas um fundo é selecionado aleatoriamente.

a. Qual é a probabilidade de o indivíduo selecionado possuir cotas do fundo misto?
b. Qual é a probabilidade de o indivíduo selecionado possuir cotas em um fundo de títulos?
c. Qual é a probabilidade de o indivíduo selecionado não possuir cotas em um fundo de ações?

12. Selecione aleatoriamente um estudante em uma determinada universidade e represente por A o evento de ele possuir um cartão de crédito Visa e por B o evento análogo para um MasterCard. Suponha que $P(A) = 0{,}5$, $P(B) = 0{,}4$ e $P(A \cap B) = 0{,}25$.

a. Calcule a probabilidade de que o indivíduo selecionado tenha pelo menos um dos dois tipos de cartão (ou seja, a probabilidade do evento $A \cup B$).
b. Qual é a probabilidade de o indivíduo selecionado não ter nenhum dos tipos de cartão?
c. Descreva, em termos de A e B, o evento em que o estudante selecionado possui um cartão Visa, mas não um MasterCard. Calcule a probabilidade desse evento.

13. Uma empresa de consultoria em informática apresenta suas propostas de três projetos. Represente por $A_i = \{\text{projeto } i \text{ fechado}\}$, para $i = 1, 2, 3$ e suponha que $P(A_1) = 0{,}22$, $P(A_2) = 0{,}25$, $P(A_3) = 0{,}28$, $P(A_1 \cap A_2) = 0{,}11$, $P(A_1 \cap A_3) = 0{,}05$, $P(A_2 \cap A_3) = 0{,}07$, $P(A_1 \cap A_2 \cap A_3) = 0{,}01$. Expresse em palavras cada evento a seguir e calcule sua probabilidade:

a. $A_1 \cup A_2$ 0,36
b. $A_1' \cap A_2'$ [*Sugestão:* $(A_1 \cup A_2)' = A_1' \cap A_2'$] = 0,64
c. $A_1 \cup A_2 \cup A_3$
d. $A_1' \cap A_2' \cap A_3'$
e. $A_1' \cap A_2' \cap A_3$
f. $(A_1' \cap A_2') \cup A_3$

14. Uma empresa de eletricidade oferece uma taxa vitalícia de energia a qualquer lar cuja utilização de energia esteja abaixo de 240 kWh durante um determinado mês. Represente por A o evento de um lar selecionado aleatoriamente em uma comunidade que não excede a utilização da taxa vitalícia em janeiro e por B o evento análogo para o mês de julho (A e B se referem ao mesmo lar). Suponha que $P(A) = 0{,}8$, $P(B) = 0{,}7$ e $P(A \cup B) = 0{,}9$. Calcule:

a. $P(A \cap B)$.
b. A probabilidade de a quantia da taxa vitalícia ser excedida em exatamente um dos dois meses. Descreva esse evento em termos de A e B.

15. Considere o tipo de secadora de roupas (a gás ou elétricas) comprada por cinco clientes diferentes em uma loja.

a. Se a probabilidade de no máximo um desses clientes fazer uma compra de uma secadora elétrica for 0,428, qual será a probabilidade de ao menos dois clientes comprarem uma secadora elétrica?
b. Se P(os cinco comprarem a gás) = 0,116 e P(os cinco comprarem elétricas) = 0,005, qual será a probabilidade de haver uma compra de ao menos uma de cada tipo?

16. Um indivíduo recebe três copos de refrigerante diferentes rotulados de C, D e P. Solicita-se que ele experimente os três e os relacione em ordem de preferência. Suponha que o mesmo refrigerante tenha sido colocado nos três copos.

a. Quais são os eventos simples nesse experimento de classificação e qual probabilidade você atribuiria a cada um deles?
b. Qual é a probabilidade de C ser classificado como o primeiro?
c. Qual é a probabilidade de C ser classificado como o primeiro e D como o último?

17. Represente por A o evento de que a próxima solicitação de assistência de um consultor de um software estatístico seja relacionada ao pacote SPSS e por B o evento

de a próxima solicitação de ajuda ser relacionada ao pacote SAS. Suponha que $P(A) = 0{,}30$ e $P(B) = 0{,}50$.

a. Por que não é o caso de $P(A) + P(B) = 1$?
b. Calcule $P(A')$.
c. Calcule $P(A \cup B)$.
d. Calcule $P(A' \cap B')$.

18. Uma caixa contém quatro lâmpadas de 40-W, cinco de 60-W e seis de 75-W. Se as lâmpadas forem selecionadas uma a uma em ordem aleatória, qual é a probabilidade de ao menos duas serem selecionadas para obter uma de 75 W?

19. A inspeção visual de juntas de solda em placas de circuitos impressos pode ser bastante subjetiva. Parte do problema se origina dos diversos tipos de defeitos de soldas (por exemplo, falta solda em pontos variados) e até da quantidade de um ou mais desses defeitos. Conseqüentemente, até mesmo inspetores altamente treinados podem discordar sobre a disposição de uma junta. Em um lote de 10.000 juntas, o inspetor *A* encontrou 724 que julgou defeituosas, o inspetor *B* encontrou 751 e 1159 foram encontradas por ao menos um dos inspetores. Suponha que uma dessas 10.000 juntas seja selecionada aleatoriamente.

a. Qual é a probabilidade de que a junta selecionada não seja julgada defeituosa por nenhum dos dois inspetores?
b. Qual é a probabilidade de que a junta selecionada seja julgada defeituosa pelo inspetor *B*, mas não pelo inspetor *A*?

20. Uma determinada fábrica opera em três turnos diferentes. No ano anterior, ocorreram 200 acidentes na fábrica. Alguns deles podem ser atribuídos em parte a condições de trabalho inseguras, enquanto os outros não estão relacionados a condições de trabalho. A tabela a seguir fornece as porcentagens de acidentes que se encaixam em cada categoria de turno de trabalho.

		Condições Inseguras	Não relacionado às condições
	Dia	10%	35%
Turno	*Alternado*	8%	20%
	Noite	5%	22%

Suponha que um dos 200 relatórios de acidente seja selecionado aleatoriamente de um arquivo de relatórios e sejam determinados o tipo de acidente e o turno.

a. Quais são os eventos simples?
b. Qual é a probabilidade de que o acidente selecionado seja atribuído a condições inseguras?
c. Qual é a probabilidade de que o acidente selecionado não tenha ocorrido no turno do dia?

21. Uma empresa de seguros oferece quatro níveis de dedução – nenhum, baixo, médio e alto – para os possuidores de apólices de seguros residenciais e três níveis diferentes – baixo, médio e alto – para os possuidores de apólices de seguros de automóveis. A tabela a seguir fornece as proporções das diversas categorias de segurados que possuem ambos os tipos de seguros. Por exemplo: a proporção de indivíduos com baixa dedução de seguro residencial e baixa dedução de seguro de automóvel é 0,06 (6% de todos os indivíduos).

	Residencial			
Automóvel	**N**	**B**	**M**	**A**
B	0,04	0,06	0,05	0,03
M	0,07	0,10	0,20	0,10
A	0,02	0,03	0,15	0,15

Suponha que um indivíduo que possua ambos os tipos de apólices seja selecionado aleatoriamente.

a. Qual é a probabilidade de que o indivíduo tenha dedução média de automóvel e alta de residência?
b. Qual é a probabilidade de que o indivíduo tenha uma dedução baixa de automóvel? Uma dedução baixa de residência?
c. Qual é a probabilidade de que um indivíduo esteja na mesma categoria para deduções de automóvel e residência?
d. Com base na resposta da parte (c), qual é a probabilidade de que as duas categorias sejam diferentes?
e. Qual é a probabilidade de que o indivíduo tenha ao menos um nível baixo de dedução?
f. Usando a resposta da parte (e), qual é a probabilidade de que nenhum nível de dedução seja baixo?

22. A rota usada por um motorista que vai ao trabalho contém dois cruzamentos com semáforos. A probabilidade de que ele tenha de parar no primeiro semáforo é 0,4, a probabilidade análoga para o segundo semáforo é 0,5 e a probabilidade de que ele tenha de parar em pelo menos um dos dois semáforos é 0,6. Qual é a probabilidade de ele ter de parar:

a. Nos dois semáforos?
b. No primeiro semáforo mas não no segundo?
c. Em exatamente um semáforo?

23. Os computadores de seis membros do corpo docente de um determinado departamento serão substituídos. Dois deles escolheram *laptops* e os outros quatro *desktops*. Suponha que apenas duas configurações podem ser feitas em um determinado dia e que os dois computadores a serem configurados sejam selecionados aleatoriamente entre os seis (implicando 15 resultados igualmente prováveis se os computadores estiverem numerados 1, 2, ..., 6, de forma que um resultado consistirá nos computadores 1 e 2, outro nos computadores 1 e 3 e assim por diante).

a. Qual é a probabilidade de que as duas configurações selecionadas sejam de *laptops*?
b. Qual é a probabilidade de que as configurações selecionadas sejam de *desktops*?
c. Qual é a probabilidade de que ao menos uma das configurações selecionadas seja de um computador *desktop*?

d. Qual é a probabilidade de que ao menos um computador de cada tipo seja escolhido para a configuração?

24. Use os axiomas para mostrar que, se um evento A estiver contido em outro evento B (ou seja, A é um subconjunto de B), então $P(A) \leq P(B)$. [*Sugestão:* neste caso, A e B, A e $B \cap A'$ são disjuntos e $B = A \cup (B \cap A')$, como pode ser visto em um diagrama de Venn.] Para os A e B genéricos, o que podemos deduzir sobre as relações $P(A \cap B)$, $P(A)$, e $P(A \cup B)$?

25. Os três principais itens opcionais de certo tipo de carro novo são transmissão automática (A), teto solar (B) e rádio com CD-player (C). Se 70% de todos os compradores solicitarem A, 80% solicitarem B, 75% solicitarem C, 85% solicitarem A ou B, 90% solicitarem A ou C, 95% solicitarem B ou C e 98% solicitarem A ou B *ou* C, calcule as probabilidades dos eventos a seguir. [*Sugestão:* "A ou B" é o evento em que no mínimo dois opcionais são solicitados.] Desenhe um diagrama de Venn e identifique todas as regiões.

a. O comprador seguinte solicita um dos três opcionais.
b. O comprador seguinte não solicita nenhum opcional.
c. O comprador seguinte solicita apenas transmissão automática e nenhum dos outros dois opcionais.
d. O comprador seguinte solicita exatamente um opcional.

26. Um determinado sistema pode ter três tipos diferentes de defeitos. Represente por A_i $(i = 1, 2, 3)$ o evento em que o sistema apresenta um defeito do tipo i. Suponha que

$P(A_1) = 0,12$ $P(A_2) = 0,07$ $P(A_3) = 0,05$
$P(A_1 \cup A_2) = 0,13$ $P(A_1 \cup A_3) = 0,14$
$P(A_2 \cup A_3) = 0,10$ $P(A_1 \cap A_2 \cap A_3) = 0,01$

a. Qual é a probabilidade de o sistema não ter um defeito do tipo 1?
b. Qual é a probabilidade de o sistema ter defeitos dos tipos 1 e 2?
c. Qual é a probabilidade de o sistema ter defeitos dos tipos 1 e 2, mas não do tipo 3?
d. Qual é a probabilidade de o sistema ter no máximo dois desses defeitos?

27. Um departamento acadêmico com cinco membros do corpo docente — Anderson, Box, Cox, Cramer e Fisher — deve selecionar dois deles para servir a um comitê de avaliação de pessoal. Como o trabalho levará muito tempo, ninguém deseja servir, de forma que foi decidido que os representantes serão selecionados por sorteio, colocando-se cinco tiras de papel em uma caixa, misturando-as e selecionando duas.

a. Qual é a probabilidade de Anderson e Box serem selecionados? (*Sugestão:* Relacione os resultados igualmente prováveis.)
b. Qual é a probabilidade de que pelo menos um dos dois membros cujos nomes sejam iniciados por C seja selecionado?
c. Se os cinco membros tiverem 3, 6, 7, 10 e 14 anos de ensino na universidade, respectivamente, qual é a probabilidade de que os dois representantes escolhidos tenham no mínimo 15 anos de ensino na universidade (tempos de ensino somados)?

28. No Exercício 5, suponha que qualquer indivíduo tenha possibilidade igualmente provável de ser designado para qualquer um dos três postos independentemente de para onde os outros sejam designados. Qual é a possibilidade de:

a. Os três membros da família serem enviados ao mesmo posto?
b. No máximo dois membros da família serem enviados ao mesmo posto?
c. Cada membro da família ser enviado a um posto diferente?

2.3 Técnicas de contagem

Quando os diversos resultados de um experimento são igualmente prováveis (a mesma probabilidade é atribuída a cada evento simples), a tarefa de calcular probabilidades se reduz a uma contagem. Em particular, se N for a quantidade de resultados de um espaço amostral e $N(A)$ for a quantidade de resultados contidos em um evento A, então

$$P(A) = \frac{N(A)}{N} \qquad (2.1)$$

Se uma lista dos resultados estiver disponível ou for de fácil construção, e N for pequeno, então o numerador e o denominador da Equação (2.1) podem ser obtidos sem o uso de quaisquer princípios de contagem geral.

Há, entretanto, muitos experimentos para os quais o esforço despendido na elaboração de tal lista é proibitivo porque N é muito grande. Explorando algumas regras gerais de contagem, é possível calcular

probabilidades da forma (2.1) sem relacionar os resultados. Essas regras também são úteis em vários problemas que envolvem resultados que não sejam igualmente prováveis. Várias das regras desenvolvidas aqui serão usadas no estudo das distribuições de probabilidades no próximo capítulo.

A regra do produto para pares ordenados

Nossa primeira regra se aplica a qualquer situação em que um conjunto (evento) consiste em pares ordenados de objetos e desejamos contar o número desses pares. Entendemos por par ordenado que, se O_1 e O_2 forem objetos, o par (O_1, O_2) será diferente do par (O_2, O_1). Por exemplo: se um indivíduo seleciona uma linha aérea para uma viagem de Los Angeles a Chicago e (após fazer negócios em Chicago) uma segunda linha para seguir até Nova York, uma possibilidade é (American, United), outra é (United, American) e outra, (United, United).

PROPOSIÇÃO

Se o primeiro elemento ou objeto de um par ordenado puder ser selecionado de n_1 formas e para cada uma dessas n_1 formas, o segundo elemento do par pode ser selecionado de n_2 formas, o que faz com que o número de pares seja n_1n_2.

Exemplo 2.17

O proprietário de uma casa em reforma solicita os serviços de um encanador e de um eletricista. Se houver 12 empreiteiros de encanamento e 9 empreiteiros eletricistas disponíveis na área, de quantas formas os empreiteiros podem ser contratados? Se representarmos os encanadores por $P_1 \ldots P_{12}$ e os eletricistas por $Q_1 \ldots Q_9$, desejamos que os pares sejam da forma (P_i, Q_j). Com $n_1 = 12$ e $n_2 = 9$, a regra do produto fornece $N = (12)(9) = 108$ formas possíveis de escolher dois tipos de empreiteiros. ■

No Exemplo 2.17, a escolha do segundo elemento do par não depende de o primeiro ter sido escolhido ou ter ocorrido. Enquanto houver o mesmo número de escolhas do segundo elemento para cada primeiro, a regra do produto será válida, mesmo quando o conjunto de segundos elementos possíveis depender do primeiro.

Exemplo 2.18

Uma família se mudou para uma cidade e precisa dos serviços de um obstetra e de um pediatra. Há duas clínicas de fácil acesso e cada uma tem dois obstetras e três pediatras. A família obterá benefícios máximos de seguro de saúde se escolher uma clínica e selecionar os dois especialistas. De quantas formas isso pode ser feito? Represente os obstetras por O_1, O_2, O_3, e O_4 e os pediatras por $P_1, \ldots P_6$. Desejamos então determinar o número de pares (O_i, P_j) para os quais O_i e P_j estão associados à mesma clínica. Como há quatro obstetras, $n_1 = 4$ e para cada um há três escolhas de pediatra, $n_2 = 3$, a aplicação da regra do produto fornece $N = n_1n_2 = 12$ escolhas possíveis. ■

Diagramas de árvore

Em diversos problemas de contagem e probabilidade, uma configuração denominada *diagrama de árvore* pode ser usada para representar graficamente todas as possibilidades. O diagrama de árvore associado ao Exemplo 2.18 é exibido na Figura 2.7. Partindo de um ponto do lado esquerdo do diagrama, para cada primeiro elemento possível de um par origina-se um segmento de reta direcionado para a direita. Cada uma dessas retas é denominada ramo de primeira geração. Para cada ramo de primeira geração criamos outro segmento de reta originando-se na extremidade do ramo, para cada escolha possível do segundo elemento do par. Cada segmento de reta é um ramo de segunda geração. Como há quatro obstetras, há quatro ramos de primeira geração e três pediatras para cada obstetra resultam em três ramos de segunda geração, originando-se de cada ramo de primeira geração.

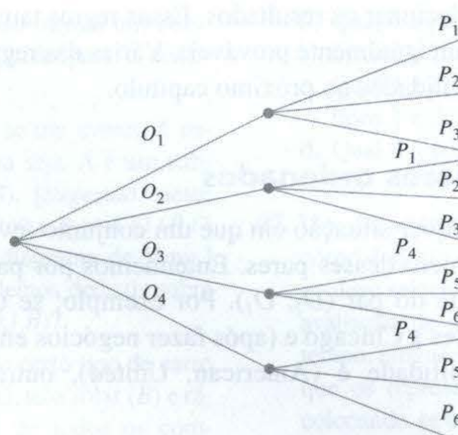


Figura 2.7 Diagrama de árvore do Exemplo 2.18

Suponha que, genericamente, haja n_1 ramos de primeira geração e para cada um deles haja n_2 ramos de segunda geração. O número total de ramos de segunda geração será então n_1n_2. Já que a extremidade de cada ramo de segunda geração corresponde a exatamente um par possível (escolher um primeiro elemento e então um segundo nos coloca exatamente na extremidade de um ramo de segunda geração), há pares n_1n_2, o que comprova a regra do produto.

A construção de um diagrama de árvore não depende de ter o mesmo número de ramos de segunda geração originando-se de cada ramo de primeira geração. Se a segunda clínica tivesse quatro pediatras, haveria apenas três ramos partindo de dois dos ramos de primeira geração e quatro partindo de cada um dos dois outros ramos de primeira geração. Um diagrama de árvore pode então ser usado para representar graficamente experimentos diferentes daqueles a que se aplica a regra do produto.

Uma regra do produto mais abrangente

Se um dado de seis lados é lançado cinco vezes consecutivas em vez de apenas duas, cada resultado possível é um conjunto de cinco números como (1, 3, 1, 2, 4) ou (6, 5, 2, 2, 2). Denominaremos um conjunto ordenado de k objetos uma ***k*-tupla** (assim um par é uma 2-tupla e um trio é uma 3-tupla). Cada resultado de experimento de lançamento de dados é uma 5-tupla.

Regra do produto para *k*-tuplas

Suponha que um conjunto consista em conjuntos ordenados de k elementos (k-tuplas) e que haja n_1 escolhas possíveis para o primeiro elemento. Para cada escolha do primeiro elemento, há n_2 escolhas possíveis do segundo elemento . . . ; para cada escolha possível dos primeiros $k - 1$ elemento, há n_k escolhas do k^o elemento. Então há $n_1 n_2 \cdot \cdots \cdot n_k$ resultados possíveis k-tuplas.

Essa regra mais geral também pode ser ilustrada por um diagrama de árvore. Apenas construa um diagrama mais elaborado, adicionando ramos de terceira geração partindo de cada extremidade de um ramo de segunda geração e assim por diante, até que os ramos de k^o geração sejam adicionados.

Exemplo 2.19 (continuação do Exemplo 2.17)

Suponha que o trabalho de reforma da casa envolva a compra de diversos utensílios de cozinha. Eles serão comprados do mesmo fornecedor e há cinco deles disponíveis na área. Com os fornecedores representados por D_1, ..., D_5, há $N = n_1n_2n_3 = (5)(12)(9) = 540$ 3-tuplas da forma (D_i, P_j, Q_k), havendo 540 formas de escolher primeiro um fornecedor de eletrodomésticos, depois um encanador e por fim um eletricista. ■

Exemplo 2.20 (continuação do Exemplo 2.18)

Se cada clínica tiver três especialistas em medicina interna e dois cirurgiões gerais, haverá $n_1n_2n_3n_4 = (4)(3)(3)(2) = 72$ formas de selecionar um médico de cada tipo, de maneira que todos os médicos sejam da mesma clínica. ■

Arranjos e Permutações

Até agora foram selecionados os elementos sucessivos de uma k-tupla a partir de conjuntos totalmente diferentes (por exemplo: fornecedores de utensílios de cozinha, encanadores e eletricistas). Em diversos lançamentos de um dado, o conjunto a partir do qual os elementos sucessivos são escolhidos é sempre $\{1, 2, 3, 4, 5, 6\}$, mas as escolhas são feitas "com reposição", de forma que o mesmo elemento pode aparecer mais de uma vez. Consideremos um conjunto determinado que consiste em n elementos diferentes e suponhamos que um k-tupla seja formado pela seleção sucessiva a partir do conjunto *sem reposição,* de modo que um elemento possa aparecer uma única vez nas k posições possíveis.

DEFINIÇÃO Qualquer seqüência ordenada de k objetos selecionados de um conjunto de n objetos distintos é denominada **arranjo** de tamanho k dos objetos. O número de arranjos de tamanho k que pode ser criado a partir dos n objetos é representado por $A_{k,n}$.

O número de arranjos de tamanho k é obtido imediatamente pela regra geral do produto. O primeiro elemento pode ser escolhido de n formas diversas; para cada uma das n formas, o segundo elemento pode ser escolhido de $n - 1$ formas e assim por diante. Finalmente, para cada forma de seleção dos primeiros $k - 1$ elementos, o k-ésimo elemento pode ser escolhido de $n - (k - 1) = n - k + 1$ formas. Então

$$A_{k,n} = n(n-1)(n-2) \cdot \cdots \cdot (n-k+2)(n-k+1)$$

Exemplo 2.21

Há 10 assistentes de professores disponíveis para correção de provas em um determinado curso. O primeiro exame consiste em quatro questões e o professor deseja selecionar um assistente diferente para corrigir cada uma (apenas um assistente por questão). De quantas formas diferentes os assistentes podem ser escolhidos para a correção? Aqui, n = número de assistentes = 10 e k = número de questões = 4. O número de atribuições diferentes de correção então será $A_{4,10} = (10)(9)(8)(7) = 5040$. ■

A utilização da notação fatorial permite que $A_{k,n}$ seja expresso de forma mais compacta.

DEFINIÇÃO Para qualquer inteiro positivo m, $m!$ é lido "fatorial de m" e definido por $m! = m(m-1) \cdot \cdots \cdot (2)(1)$. Além disso, $0! = 1$.

Usando a notação fatorial, $(10)(9)(8)(7) = (10)(9)(8)(7)(6!)/6! = 10!/6!$. De forma mais geral,

$$A_{k,n} = n(n-1) \cdot \cdots \cdot (n-k+1)$$

$$= \frac{n(n-1) \cdot \cdots \cdot (n-k+1)(n-k)(n-k-1) \cdot \cdots \cdot (2)(1)}{(n-k)(n-k-1) \cdot \cdots \cdot (2)(1)}$$

que se transforma em

$$P_{k,n} = \frac{n!}{(n-k)!}$$

Por exemplo: $A_{3,9} = 9!/(9-3)! = 9!/6! = 9 \cdot 8 \cdot 7 \cdot 6!/6! = 9 \cdot 8 \cdot 7$. Observe também que, como $0! = 1$, $0! = 1$, $P_{n,n} = n!/(n-n)! = n!/0! = n!/1 = n!$, conforme mencionado.

Combinações

Há muitos problemas de contagem em que é fornecido um conjunto de n objetos distintos e se deseja contar o número de subconjuntos não-ordenados de tamanho k. Por exemplo: no *bridge* importam apenas as 13 cartas da mão e não a ordem em que elas são dadas. Na formação de um comitê, a ordem em que os membros são relacionados normalmente não tem importância.

DEFINIÇÃO Dado um conjunto de n objetos diferentes, qualquer subconjunto não-ordenado de tamanho k é denominado **combinação.** O número de combinações de tamanho k que podem ser formadas a partir de n objetos distintos é representado por $\binom{n}{k}$. (Essa notação é mais comum na probabilidade do que $C_{k,n}$, que seria análogo à notação de permutações.)

O número de combinações de tamanho k de um determinado conjunto é menor que o número de arranjos porque, quando a ordem não importa, diversos arranjos correspondem à mesma combinação. Considere, por exemplo, o conjunto $\{A, B, C, D, E\}$, consistindo em cinco elementos. Há $5!/(5 - 3)! = 60$ arranjos de tamanho 3. Há seis arranjos de tamanho 3 que consistem nos elementos A, B e C, já que esses três elementos podem ser ordenados de $3 \cdot 2 \cdot 1 = 3! = 6$ formas: (A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B) e (C, B, A). Esses seis arranjos são equivalentes a uma única combinação $\{A, B, C\}$. De forma similar, para qualquer outra combinação de tamanho 3, há 3! arranjos a cada uma obtida pela ordenação dos três objetos. Dessa forma,

$$60 = A_{3,5} = \binom{5}{3} 3!; \quad \text{então} \binom{5}{3} = \frac{60}{3!} = 10$$

Essas dez combinações são

$\{A, B, C\}$ $\{A, B, D\}$ $\{A, B, E\}$ $\{A, C, D\}$ $\{A, C, E\}$ $\{A, D, E\}$, $\{B, C, D\}$
$\{B, C, E\}$ $\{B, D, E\}$ $\{C, D, E\}$

Quando há n objetos diferentes, qualquer arranjo de tamanho k é obtido pela ordenação dos k objetos não-ordenados de uma combinação em uma de $k!$ formas, de modo que a quantidade de arranjos seja o produto de $k!$ e do número de combinações. Isso fornece

$$\binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!}$$

Observe que $\binom{n}{n} = 1$ e $\binom{n}{0} = 1$, já que há apenas uma forma de escolhermos um conjunto de (todos) n elementos ou de nenhum elemento, e $\binom{n}{1} = n$, já que há n subconjuntos de tamanho 1.

Exemplo 2.22

Uma mão de *bridge* consiste em quaisquer 13 cartas selecionadas de um baralho de 52 cartas sem importar a ordem. Há $\binom{52}{13} = 52!/13!39!$ diferentes mãos de *bridge*, o que dá cerca de 635 bilhões. Como há 13 cartas em cada naipe, o número de mãos formadas unicamente por paus e/ou espadas (sem cartas vermelhas) é $\binom{26}{13} = 26!/13!13! = 10{,}400{,}600$. Uma dessas $\binom{26}{13}$ mãos é formada totalmente por espadas e outra por paus, de forma que há $[\binom{26}{13} - 2]$ mãos formadas totalmente por espadas e paus na mão. Suponha que uma mão de *bridge* seja dada de um baralho bem embaralhado (ou seja, 13 cartas selecionadas aleatoriamente entre as 52 possibilidades) e seja

A = {a mão formada inteiramente por espadas e paus com os dois naipes}
B = {a mão formada por exatamente dois naipes}.

Os $N = \binom{52}{13}$ resultados possíveis são igualmente prováveis, de forma que

$$P(A) = \frac{N(A)}{N} = \frac{\binom{26}{13} - 2}{\binom{52}{13}} = 0{,}0000164$$

Como há $\binom{4}{2} = 6$ combinações formadas por dois naipes, das quais espadas e paus são tal combinação,

$$P(B) = \frac{6\left[\binom{26}{13} - 2\right]}{\binom{52}{13}} = 0{,}0000983$$

Ou seja, uma mão totalmente formada por cartas de exatamente dois dos quatro naipes ocorre cerca de uma vez a cada 10.000. Se você joga *bridge* apenas uma vez por mês, é provável que nunca receba uma dessas mãos. ■

Exemplo 2.23

Um armazém de universidade recebeu uma entrega de 25 impressoras, das quais 10 são impressoras a laser e 15 são a jato de tinta. Se 6 das 25 forem selecionadas aleatoriamente para serem verificadas por um técnico, qual será a probabilidade de que exatamente 3 delas sejam a laser (sendo as outras 3 a jato de tinta)?

Represente por D_3 = {exatamente 3 das 6 selecionadas são impressoras a jato de tinta}. Assumindo que qualquer conjunto específico de 6 impressoras tem a mesma chance de ser escolhido em relação a qualquer outro conjunto de 6, temos resultados igualmente prováveis, de forma que $P(D_3) = N(D_3)/N$, onde N é o número de formas de escolha das 6 impressoras dentre as 25 e $N(D_3)$ é o número de formas de seleção de 3 impressoras a laser e 3 a jato de tinta. Assim, $N = \binom{25}{6}$. Para obter $N(D_3)$, pense em primeiro escolher 3 dos 15 modelos a jato de tinta e então 3 impressoras a laser. Há $\binom{15}{3}$ formas de escolha de 3 modelos a jato de tinta e $\binom{10}{3}$ formas de escolha de 3 impressoras a laser. Nesse caso, $N(D_3)$ é o produto desses dois números (visualize um diagrama de árvore – estamos realmente usando a regra do produto), de forma que

$$P(D_3) = \frac{N(D_3)}{N} = \frac{\binom{15}{3}\binom{10}{3}}{\binom{25}{6}} = \frac{\frac{15!}{3!12!} \cdot \frac{10!}{3!7!}}{\frac{25!}{6!19!}} = 0{,}3083$$

Represente por D_4 = {exatamente 4 das 6 impressoras selecionadas são modelos a jato de tinta} e defina D_5 e D_6 de forma análoga. Então a probabilidade de ao menos 3 impressoras a jato de tinta serem selecionadas é

$$P(D_3 \cup D_4 \cup D_5 \cup D_6) = P(D_3) + P(D_4) + P(D_5) + P(D_6)$$

$$= \frac{\binom{15}{3}\binom{10}{3}}{\binom{25}{6}} + \frac{\binom{15}{4}\binom{10}{2}}{\binom{25}{6}} + \frac{\binom{15}{5}\binom{10}{1}}{\binom{25}{6}} + \frac{\binom{15}{6}\binom{10}{0}}{\binom{25}{6}} = 0{,}8530$$ ■

Exercícios | Seção 2.3 (29–44)

29. O conselho de estudantes de engenharia de certa faculdade possui um aluno representante de cada uma das áreas de engenharia (civil, elétrica, produção de materiais e mecânica). De quantas formas é possível:

a. Selecionar um presidente e um vice-presidente?

b. Selecionar um presidente, um vice-presidente e um secretário?

c. Selecionar dois membros para o conselho da reitoria?

30. Um amigo meu vai oferecer um jantar. Sua adega inclui 8 garrafas de zinfandel, 10 de merlot e 12 de cabernet (ele só toma vinho tinto), todos de vinícolas diferentes.

a. Se ele quiser servir 3 garrafas de zinfandel e a ordem para servir for importante, de quantas formas pode fazê-lo?

b. Se forem selecionadas 6 garrafas de vinho aleatoriamente entre as 30 disponíveis para servir, quantas formas há de selecioná-las?

c. Se forem selecionadas 6 garrafas aleatoriamente, de quantas formas será possível selecionar duas garrafas de cada variedade?

d. Se forem selecionadas 6 garrafas aleatoriamente, qual a probabilidade de serem escolhidas duas garrafas de cada variedade?

e. Se forem selecionadas 6 garrafas aleatoriamente, qual a probabilidade de todas serem do mesmo tipo?

31. a. Beethoven escreveu 9 sinfonias e Mozart escreveu 27 concertos para piano. Se a estação de rádio de uma universidade desejar tocar primeiro uma sinfonia de Beethoven e depois um concerto de Mozart, de quantas formas pode ser feita a escolha?

b. O gerente da rádio decide que em cada noite sucessiva (7 dias por semana) será tocada uma sinfonia de Beethoven, seguida por um concerto de piano de Mozart, seguido por um quarteto de cordas de Schubert (de um total de 15). Por cerca de quantos anos essa política pode continuar, antes que exatamente o mesmo programa seja repetido?

32. Uma cadeia de lojas de aparelhos de som oferece um preço especial para o conjunto completo de componentes (aparelho de som, CD-*player*, alto-falantes, toca-fitas). O comprador tem a seguinte opção de fabricante para cada componente:

Aparelho de som:	Kenwood, Onkyo, Pioneer, Sony, Sherwood
CD-player:	Onkyo, Pioneer, Sony, Technics
Alto-falantes:	Boston, Infinity, Polk
Toca-fitas:	Onkyo, Sony, Teac, Technics

Um painel de ligação da loja permite ao cliente conectar qualquer seleção de componentes (um de cada tipo). Use a regra do produto para responder as seguintes perguntas:

a. De quantas formas um componente de cada tipo pode ser selecionado?

b. De quantas formas os componentes podem ser selecionados se o aparelho de som e o CD-player forem Sony?

c. De quantas formas os componentes podem ser selecionados se nenhum for Sony?

d. De quantas formas pode ser feita uma seleção se ao menos um componente Sony for incluído?

e. Se os interruptores forem ligados de forma totalmente aleatória, qual será a probabilidade de o sistema selecionado conter ao menos um componente Sony? E exatamente um componente Sony?

33. Logo após terem sido colocados em serviço, alguns ônibus fabricados por uma determinada empresa apresentaram trincas na parte inferior do chassi. Suponha que uma cidade tenha 25 desses ônibus e que haja trincas em 8 deles.

a. Há quantas formas de selecionar uma amostra de 5 ônibus dos 25 para uma inspeção completa?

b. De quantas formas uma amostra de 5 ônibus pode conter exatamente 4 com trincas visíveis?

c. Se uma amostra de 5 ônibus for selecionada ao acaso, qual é a probabilidade de exatamente 4 dos 5 apresentarem trincas visíveis?

d. Se os ônibus forem selecionados como na parte (c), qual é a probabilidade de que pelo menos 4 dos ônibus selecionados tenham trincas visíveis?

34. Uma instalação de produção emprega 20 operários no turno diurno, 15 operários no noturno e 10 operários no da madrugada. Um consultor de controle de qualidade deve selecionar 6 desses operários para entrevistas detalhadas. Suponha que a seleção seja feita de tal forma que qualquer grupo específico de 6 operários tenha a mesma possibilidade de ser selecionado que qualquer outro grupo (escolha de 6 nomes sem reposição entre 45).

a. Quantas escolhas contêm 6 operários do turno diurno? Qual é a probabilidade de os 6 operários selecionados serem do turno diurno?

b. Qual é a probabilidade de os 6 operários selecionados serem do mesmo turno?

c. Qual é a probabilidade de pelo menos dois turnos diferentes terem representantes entre os operários selecionados?

d. Qual é a probabilidade de pelo menos um dos turnos não ter representante na amostra de operários?

35. Um departamento acadêmico com cinco membros docentes restringiu sua escolha para chefe do departamento ao candidato *A* ou candidato *B*. Cada membro então votou em um dos candidatos em uma cédula.

Suponha que haja três votos para o candidato *A* e dois para o candidato *B*. Se as cédulas forem selecionadas para contagem em ordem aleatória, qual é a probabilidade de *A* se manter à frente de *B* na contagem dos votos (por exemplo: esse evento ocorre se a ordem selecionada for *AABAB*, mas não se for *ABBAA*)?

36. Um pesquisador está estudando os efeitos da temperatura, da pressão e do tipo de catalisador no resultado de certa reação química. Três temperaturas diferentes, quatro pressões diferentes e cinco catalisadores diferentes são considerados.

a. Se cada teste envolver a utilização de uma única temperatura, pressão e catalisador, quantos testes serão possíveis?

b. Quantos testes envolvem o uso da menor temperatura e de duas menores pressões?

37. Refira-se ao o exercício 36 e suponha que cinco testes diferentes sejam feitos no primeiro dia de experiência. Se os cinco forem selecionados aleatoriamente dentre todas as possibilidades, de forma que qualquer grupo de cinco tenha a mesma probabilidade de escolha, qual a probabilidade de que um catalisador diferente seja usado em cada teste?

38. Uma caixa em um depósito contém quatro lâmpadas de 40W, cinco de 60W e seis de 75W. Suponha que três lâmpadas sejam selecionadas aleatoriamente.

a. Qual a probabilidade de que exatamente duas das lâmpadas selecionadas sejam de 75W?

b. Qual a probabilidade de que as três lâmpadas selecionadas tenham a mesma potência?

c. Qual a probabilidade de que uma lâmpada de cada tipo seja selecionada?

d. Suponha que as lâmpadas sejam selecionadas uma a uma até que seja encontrada uma de 75W. Qual a probabilidade de que seja necessário examinar pelo menos seis lâmpadas?

39. Foram recebidos quinze telefones em uma assistência técnica autorizada. Cinco deles são celulares, cinco são sem fio e os outros cinco são telefones com fio. Suponha que esses componentes sejam numerados aleatoriamente de 1, 2, ... , a 15 para definir a ordem em que serão consertados.

a. Qual é a probabilidade de que todos os telefones sem fio estejam entre os primeiros 10 a serem consertados?

b. Qual é a probabilidade de que após consertar 10 telefones, somente dois dos três tipos ainda estejam com conserto pendente?

c. Qual é a probabilidade de que dois telefones de cada tipo estejam entre os primeiros seis a serem consertados?

40. Três moléculas do tipo *A*, três do tipo *B*, três do tipo *C* e três do tipo *D* serão vinculadas uma à outra para formar uma cadeia molecular. Uma molécula desse tipo é *ABCDABCDABCD* e outra é *BCDDAAABDBCC*.

a. Quantas moléculas desse tipo podem ser formadas? (*Sugestão:* se as três moléculas *A* forem diferentes uma da outra $-A_1, A_2$ e A_3 – assim como as *B*, *C* e *D*, quantas moléculas haverá? Como esse número será reduzido se não houver distinção entre as moléculas *A*?)

b. Suponha que uma molécula composta do tipo descrito seja selecionada aleatoriamente. Qual é a probabilidade de que todas as moléculas de cada tipo estejam juntas (como em *BBBAAADDDCCC*)?

41. Uma professora de matemática deseja marcar uma reunião com cada um de seus oito assistentes, quatro homens e quatro mulheres, para discutir o curso de cálculo. Suponha que todas as seqüências de reuniões sejam igualmente prováveis de serem escolhidas.

a. Qual é a probabilidade de que pelo menos uma assistente mulher esteja entre as primeiras três pessoas com quem a professora se reunirá?

b. Qual é a probabilidade de que após as primeiras cinco reuniões ela tenha se reunido com todas as assistentes mulheres?

c. Suponha que a professora tenha os mesmos oito assistentes no semestre seguinte e marque reuniões sem considerar a ordem seguida no primeiro semestre. Qual a probabilidade de que a ordem dos compromissos seja diferente?

42. Três casais compraram ingressos de teatro e estão sentados em uma fileira que consiste em apenas seis assentos. Se eles se sentarem de uma forma totalmente aleatória (ordem aleatória), qual será a probabilidade de Jim e Paula (marido e mulher) se sentarem nos dois assentos da esquerda? Qual é a probabilidade de Jim e Paula se sentarem um ao lado do outro? Qual é a probabilidade de que ao menos uma das esposas se sente ao lado de seu respectivo marido?

43. No pôquer de cinco cartas, um *straight* consiste em cinco cartas com denominações adjacentes (por exemplo: 9 de paus, 10 de copas, valete de copas, dama de espadas e rei de paus). Assumindo que os ases podem ficar nas duas pontas, se você receber um jogo de cinco cartas, qual é a probabilidade de que ele será um *straight* com a carta 10 alta? Qual é a probabilidade de se receber um *straight*? Qual é a probabilidade de ser um *straight flush* (todas as cartas do mesmo naipe)?

44. Mostre que $\binom{n}{k} = \binom{n}{n-k}$. Forneça uma interpretação envolvendo subconjuntos.

2.4 Probabilidade Condicional

As probabilidades atribuídas a diversos eventos dependem do que seja conhecido sobre a situação experimental quando a atribuição é feita. Depois da atribuição inicial, informações parciais sobre ou relevantes para o resultado do experimento podem estar disponíveis. Essas informações podem fazer com que revisemos algumas atribuições de probabilidade. Para um determinado evento A, usamos $P(A)$ para representar a probabilidade atribuída a A. Agora interpretamos $P(A)$ como a probabilidade original ou incondicional do evento A. Nesta seção, examinamos como a informação "ocorreu um evento B" afeta a probabilidade atribuída a A. Por exemplo: A pode se referir a um indivíduo que tenha determinada doença pela presença de alguns sintomas. Se for feito um exame de sangue no indivíduo e o resultado for negativo (B = exame de sangue negativo), então a probabilidade de ter a doença será alterada (ela deve diminuir, mas normalmente não para zero, já que os exames de sangue não são infalíveis). Usaremos a notação $P(A \mid B)$ para representar a **probabilidade condicional de A dado que ocorreu o evento B.**

Exemplo 2.24

Componentes complexos são montados em uma fábrica que usa duas linhas de montagem diferentes: A e A'. A linha A usa equipamentos mais antigos que A', de forma que é mais lenta e um pouco menos confiável. Suponha que em determinado dia, a linha A tenha montado 8 componentes, dos quais 2 foram identificados como defeituosos (B) e 6 como não-defeituosos (B'), ao passo que a linha A' produziu 1 defeituoso e 9 não-defeituosos. Essas informações estão resumidas na tabela a seguir.

		Condição	
		B	**B'**
Linha	**A**	2	6
	A'	1	9

Não tendo essas informações, o gerente de vendas seleciona aleatoriamente 1 desses 18 componentes para uma demonstração. Antes da demonstração

$$P(\text{componente selecionado da linha } A) = P(A) = \frac{N(A)}{N} = \frac{8}{18} = 0{,}44$$

Entretanto, se o componente escolhido tiver defeito, o evento B terá ocorrido, de forma que o componente deve ser 1 dos 3 na coluna B da tabela. Como esses 3 componentes são igualmente prováveis entre si após a ocorrência de B,

$$P(A \mid B) = \frac{2}{3} = \frac{\frac{2}{18}}{\frac{3}{18}} = \frac{P(A \cap B)}{P(B)} \qquad (2.2)$$

■

Na Equação (2.2), a probabilidade condicional é expressa como uma relação entre probabilidades não-condicionais: o numerador é a probabilidade de interseção dos dois eventos, enquanto o denominador é a probabilidade do evento de condição B. Um diagrama de Venn esclarece a relação (veja a Figura 2.8).

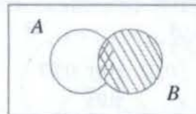


Figura 2.8 Demonstração da definição de probabilidade condicional

Dado que B ocorreu, o espaço amostral relevante não é mais $\mathcal{S}$, mas consiste em resultados contidos em B; A ocorreu se, e somente se, um dos resultados da interseção tiver ocorrido, de forma que a probabilidade condicional de A dado B é proporcional a $P(A \cap B)$. A proporção constante $1/P(B)$ é utilizada para assegurar que a probabilidade $P(B \mid B)$ do novo espaço amostral de B seja igual a 1.

Definição de probabilidade condicional

O Exemplo 2.24 demonstra que, quando os resultados são igualmente prováveis, o cálculo de probabilidades condicionais pode se basear na intuição. Quando os experimentos são mais complicados, a intuição pode nos enganar, portanto queremos uma definição geral de probabilidade condicional que forneça respostas intuitivas em problemas simples. O diagrama de Venn e a Equação (2.2) indicam a definição apropriada.

DEFINIÇÃO

Para quaisquer dois eventos A e B com $P(B) > 0$, a **probabilidade condicional de A dado que ocorreu B** é definida por

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{2.3}$$

Exemplo 2.25

Suponha que, de todos os indivíduos que compram uma determinada câmera digital, 60% incluem um cartão de memória opcional na compra, 40% incluem uma pilha extra e 30% incluem um cartão e uma pilha. Considere a seleção aleatória de um comprador e sejam A = {compra de cartão de memória} e B = {compra de pilha}. Dessa forma, $P(A) = 0{,}60$, $P(B) = 0{,}40$ e $P(\text{compra de ambos}) = P(A \cap B) = 0{,}30$. Dado que o indivíduo selecionado comprou uma pilha extra, a probabilidade de compra de um cartão opcional é

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0{,}30}{0{,}40} = 0{,}75$$

Isto é, de todos os que compraram uma pilha extra, 75% compraram um cartão de memória extra. De forma análoga,

$$P(\text{pilha} \mid \text{cartão de memória}) = P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0{,}30}{0{,}60} = 0{,}50$$

Observe que $P(A \mid B) \neq P(A)$ e $P(B \mid A) \neq P(B)$. ■

Exemplo 2.26

Uma revista publica três colunas, intituladas "Arte" (A), "Livros" (B) e "Cinema" (C). Os hábitos de leitura de um leitor selecionado aleatoriamente em relação a essas colunas são:

Lê regularmente	A	B	C	$A \cap B$	$A \cap C$	$B \cap C$	$A \cap B \cap C$
Probabilidade	0,14	0,23	0,37	0,08	0,09	0,13	0,05

(Veja a Figura 2.9.)

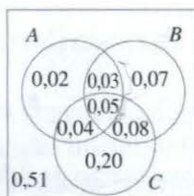


Figura 2.9 Diagrama de Venn do Exemplo 2.26

Assim, temos:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0{,}08}{0{,}23} = 0{,}348$$

$$P(A \mid B \cup C) = \frac{P(A \cap (B \cup C))}{P(B \cup C)} = \frac{0{,}04 + 0{,}05 + 0{,}03}{0{,}47} = \frac{0{,}12}{0{,}47} = 0{,}255$$

$$P(A \mid \text{lê pelo menos uma}) = P(A \mid A \cup B \cup C) = \frac{P(A \cap (A \cup B \cup C))}{P(A \cup B \cup C)}$$

$$= \frac{P(A)}{P(A \cup B \cup C)} = \frac{0{,}14}{0{,}49} = 0{,}286$$

e

$$P(A \cup B \mid C) = \frac{P((A \cup B) \cap C)}{P(C)} = \frac{0{,}04 + 0{,}05 + 0{,}08}{0{,}37} = 0{,}459$$

■

A regra da multiplicação para $P(A \cap B)$

A definição de probabilidade condicional fornece o resultado a seguir, obtido pela multiplicação de ambos os lados da equação (2.3) por *P(B)*.

> A Regra da multiplicação
>
> $$P(A \cap B) = P(A \mid B) \cdot P(B)$$

Essa regra é importante porque freqüentemente se deseja $P(A \cap B)$, ao passo que $P(B)$ e $P(A \mid B)$ podem ser especificados pela descrição do problema. Considerando a expressão $P(B \mid A)$ obtém-se $P(A \cap B) = P(B \mid A) \cdot P(A)$.

Exemplo 2.27

Quatro indivíduos responderam a uma solicitação de um banco de sangue para doação. Nenhum deles doou sangue antes, de forma que seus tipos sangüíneos são desconhecidos. Suponha que apenas o tipo O+ seja desejado e apenas um dos quatro indivíduos tenha esse tipo sangüíneo. Se os doadores potenciais forem selecionados em ordem aleatória para determinação do tipo sangüíneo, qual será a probabilidade de que pelo menos três indivíduos tenham de ser testados para obtenção do tipo desejado?

Fazendo B = {primeiro tipo não O+} e A = {segundo tipo não O+}, $P(B) = \frac{3}{4}$. Dado que o primeiro tipo não é O+, dois dos três indivíduos restantes não serão O+, de forma que $P(A \mid B) = \frac{2}{3}$. A regra de multiplicação agora fornece:

$$P(\text{pelo menos três indivíduos foram testados}) = P(A \cap B) = P(A \mid B) \cdot P(B) = \frac{2}{3} \cdot \frac{3}{4} = \frac{6}{12} = 0{,}5$$ ■

A regra da multiplicação é mais útil quando o experimento consiste em diversas etapas consecutivas. O evento condicional B então descreve o resultado da primeira etapa e A descreve o resultado da segunda, de forma que $P(A \mid B)$ – condicionado ao que ocorre primeiro – normalmente é conhecida. A regra é facilmente estendida a experimentos que envolvem mais de duas etapas. Por exemplo:

$$P(A_1 \cap A_2 \cap A_3) = P(A_3 \mid A_1 \cap A_2) \cdot P(A_1 \cap A_2) = P(A_3 \mid A_1 \cap A_2) \cdot P(A_2 \mid A_1) \cdot P(A_1) \quad (2.4)$$

onde A_1 ocorre primeiro, seguido por A_2 e finalmente por A_3.

Exemplo 2.28

Para o experimento de definição de tipo sangüíneo do Exemplo 2.27,

$$P(\text{terceiro tipo é O+}) = P(\text{terceiro é} \mid \text{primeiro não é} \cap \text{segundo não é}) \cdot P(\text{segundo não é} \mid \text{primeiro não é}) \cdot P(\text{primeiro não é}) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{4} = 0{,}25$$ ■

Quando o experimento consistir uma seqüência de diversas etapas, é conveniente representá-las em um diagrama de árvore. Depois que tivermos um diagrama de árvore apropriado, as probabilidades e as probabilidades condicionais são inseridas nos diversos ramos; o que é feito com o uso da regra da multiplicação de forma direta.

Exemplo 2.29

Uma cadeia de lojas de vídeo vende três marcas diferentes de videocassetes. Dessas vendas, 50% são da marca 1 (a mais barata), 30% são da marca 2 e 20% são da marca 3. Cada fabricante oferece um ano de garantia para peças e mão-de-obra. É sabido que 25% dos videocassetes da marca 1 necessitam de reparos de garantia, enquanto os percentuais correspondentes para as marcas 2 e 3 são 20% e 10%, respectivamente.

1. Qual é a probabilidade de que um comprador selecionado aleatoriamente compre um videocassete da marca 1 que precise de reparo durante a garantia?
2. Qual é a probabilidade de que um comprador selecionado aleatoriamente possua um aparelho que necessite de reparos durante a garantia?
3. Se um cliente voltar à loja com um videocassete que precise de reparos em garantia, qual é a probabilidade de ele ser da marca 1? E da marca 2? E da marca 3?

A primeira etapa do problema envolve um cliente que escolhe uma de três marcas de videocassete. Seja A_i = {compra da marca i}, para i = 1, 2 e 3. Então $P(A_1) = 0{,}50$, $P(A_2) = 0{,}30$ e $P(A_3) = 0{,}20$. Depois que a marca do videocassete for selecionada, a segunda etapa envolve a observação da necessidade de reparo na garantia. Com B = {precisa de reparo} e B' = {não precisa de reparo), as informações fornecidas implicam $P(B \mid A_1) = 0{,}25$, $P(B \mid A_2) = 0{,}20$, e $P(B \mid A_3) = 0{,}10$.

O diagrama de árvore que representa essa situação experimental é exibido na Figura 2.10. Os ramos iniciais correspondem a diferentes marcas de videocassetes e há ramos de segunda geração que partem da extremidade de cada ramo inicial: um para "precisa de reparo" e outro para "não precisa de reparo". A probabilidade

$P(A_i)$ é exibida no ramo inicial, enquanto as probabilidades condicionais $P(B \mid A_i)$ e $P(B' \mid A_i)$ são exibidas nos ramos de segunda geração. À direita de cada ramo de segunda geração correspondente à ocorrência de B exibimos o produto das probabilidades nos ramos que vão até o ponto. Isto é simplesmente a regra de multiplicação em ação. A resposta à pergunta proposta em 1, portanto, é $P(A_1 \cap B) = P(B \mid A_1) \cdot P(A_1) = 0{,}125$. A resposta à pergunta 2 é

$$\begin{aligned} P(B) &= P[(\text{marca 1 e reparo}) \text{ ou } (\text{marca 2 e reparo}) \text{ ou } (\text{marca 3 e reparo})] \\ &= P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \\ &= 0{,}125 + 0{,}060 + 0{,}020 = 0{,}205 \end{aligned}$$

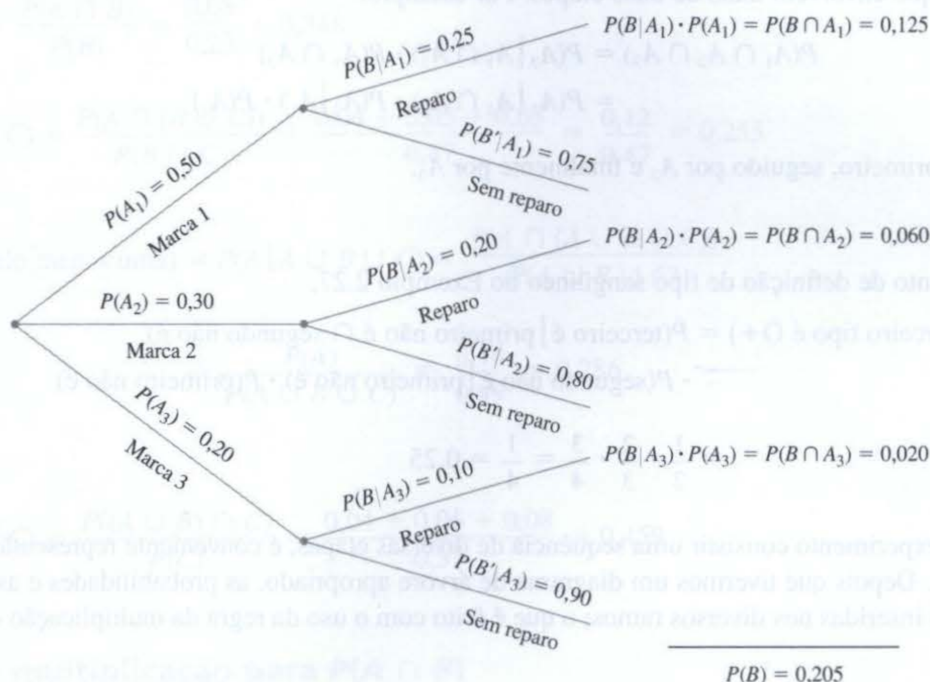


Figure 2.10 Diagrama de árvore do Exemplo 2.29

Finalmente,

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{0{,}125}{0{,}205} = 0{,}61$$

$$P(A_2 \mid B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{0{,}060}{0{,}205} = 0{,}29$$

e

$$P(A_3 \mid B) = 1 - P(A_1 \mid B) - P(A_2 \mid B) = 0{,}10$$

Observe que a *probabilidade anterior* ou inicial da marca 1 é 0,50, enquanto, depois que se sabe que o videocassete selecionado precisa de reparos, a *probabilidade posterior* da marca 1 aumenta para 0,61. Isso ocorre porque os videocassetes da marca 1 são igualmente mais prováveis de precisar de reparos do que os das outras marcas. A probabilidade posterior da marca 3 é $P(A_3 \mid B) = 0{,}10$, que é muito menor que a probabilidade anterior $P(A_3) = 0{,}20$. ■

Teorema de Bayes

O cálculo de uma probabilidade posterior $P(A_j \mid B)$ a partir de probabilidades anteriores dadas $P(A_i)$ e das probabilidades condicionais $P(B \mid A_i)$ ocupa uma posição central em teoria elementar de probabilidade. A regra geral

desses cálculos, que na verdade é uma aplicação simples da regra de multiplicação, remete ao reverendo Thomas Bayes, que viveu no século XVIII. Para expressá-la, precisamos de outro resultado. Lembre-se de que os eventos $A_1, \ldots, A_k$ são mutuamente exclusivos se nenhum par deles tiver resultados comuns. Os eventos são *exaustivos* se um A_i tiver de ocorrer para que $A_1 \cup \cdots \cup A_k = \mathcal{S}$.

Lei ou Teorema da probabilidade total

Sejam $A_1, \ldots, A_k$ eventos mutuamente exclusivos e exaustivos. Então, para qualquer outro evento B,

$$P(B) = P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_k)P(A_k) \qquad (2.5)$$
$$= \sum_{i=1}^{k} P(B \mid A_i)P(A_i)$$

Demonstração

Visto que os eventos A_i são mutuamente exclusivos e exaustivos, se B ocorrer será com exatamente um dos A_i, isto é, $B = (A_1 \text{ e } B)$ ou ... ou $(A_k \text{ e } B) = (A_1 \cap B) \cup \cdots \cup (A_k \cap B)$, onde os eventos $(A_i \cap B)$ são mutuamente exclusivos. Essa "partição de B" é ilustrada na Figura 2.11. Assim,

$$P(B) = \sum_{i=1}^{k} P(A_i \cap B) = \sum_{i=1}^{k} P(B \mid A_i)P(A_i)$$

conforme desejado.

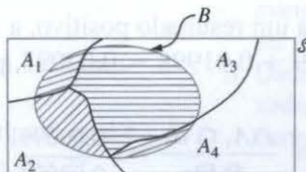


Figura 2.11 Partição de B pelos eventos mutuamente exclusivos e exaustivos A_i ■

Um exemplo do uso da Equação (2.5) foi mostrado na resposta à pergunta 2 do Exemplo 2.29, onde $A_1 = \{\text{marca } 1\}$, $A_2 = \{\text{marca } 2\}$, $A_3 = \{\text{marca } 3\}$ e $B = \{\text{reparo}\}$.

Teorema de Bayes

Sejam $A_1, A_2, \ldots, A_k$ uma coleção de k eventos mutuamente exclusivos e exaustivos com $P(A_i) > 0$ para $i = 1, \ldots, k$. Então, para qualquer outro evento B em que $P(B) > 0$,

$$P(A_j \mid B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^{k} P(B \mid A_i) \cdot P(A_i)} \qquad j = 1, \ldots, k \qquad (2.6)$$

A transição da segunda para a terceira expressão em (2.6) baseia-se na aplicação da regra da multiplicação no numerador e da lei da probabilidade total no denominador.

A proliferação de eventos e índices na expressão (2.6) pode intimidar usuários com pouca experiência em probabilidade. Desde que haja relativamente poucos eventos na partição, pode-se usar um gráfico de árvore

(como no Exemplo 2.29) como base para calcular as probabilidades posteriores sem referência explícita ao teorema de Bayes.

Exemplo 2.30

Incidência de doença rara. Apenas 1 em 1000 adultos é acometido por uma doença rara para a qual foi desenvolvido um teste de diagnóstico. O teste funciona de tal forma que, se o indivíduo tiver a doença, o resultado do teste será positivo em 99% das vezes e, se não a tiver, será positivo em apenas 2% das vezes. Se um indivíduo selecionado aleatoriamente for testado e o resultado for positivo, qual é a probabilidade de ele ter a doença?

Para usar o teorema de Bayes, considere por A_1 = {indivíduo tem a doença}, A_2 = {indivíduo não tem a doença} e B = {resultado do teste positivo}. Então, $P(A_1) = 0{,}001$, $P(A_2) = 0{,}999$, $P(B \mid A_1) = 0{,}99$, e $P(B \mid A_2) = 0{,}02$. O diagrama de árvore deste problema está na Figura 2.12.

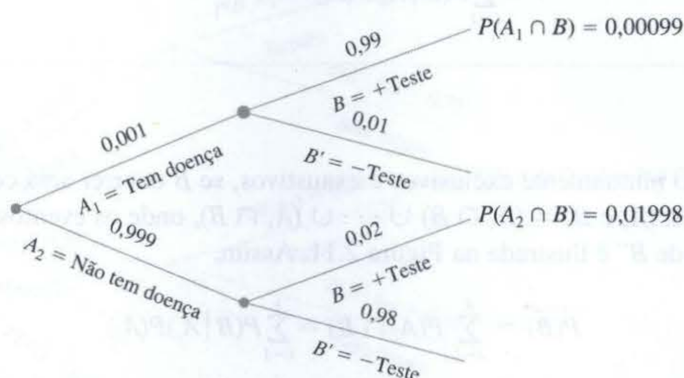


Figura 2.12 Diagrama de árvore para o problema da doença rara

Próximo a cada ramo correspondente a um resultado positivo, a regra de multiplicação indica as probabilidades registradas. Portanto, $P(B) = 0{,}00099 + 0{,}01998 = 0{,}02097$, pelo qual obtemos

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{0{,}00099}{0{,}02097} = 0{,}047$$

O resultado parece ir de encontro à intuição. O resultado do teste parece tão preciso que esperamos que alguém com um resultado positivo tenha alta probabilidade de ter a doença, enquanto a probabilidade condicional calculada é de apenas 0,047. Entretanto, como a doença é rara e o teste é moderadamente confiável, a maior parte dos resultados positivos dos testes provém de erros e não de indivíduos doentes. A probabilidade de ter a doença foi aumentada por um fator multiplicativo 47 (da probabilidade anterior 0,001 para a posterior 0,047); mas, para se obter maior aumento na probabilidade posterior, é necessário um teste de diagnóstico com menores taxas de erros. Se a doença não fosse tão rara (por exemplo, 25% de incidência na população), as taxas de erros do presente teste forneceriam um bom diagnóstico. ■

Exercícios | Seção 2.4 (45–67)

45. A população de um país consiste em três grupos étnicos. Cada indivíduo pertence a um de quatro grupos sangüíneos principais. A *tabela de probabilidade conjunta* fornece as proporções de indivíduos nas diversas combinações de grupos étnicos–grupos sangüíneos.

		Grupo sangüíneo			
		O	A	B	AB
	1	0,082	0,106	0,008	0,004
Grupo étnico	2	0,135	0,141	0,018	0,006
	3	0,215	0,200	0,065	0,020

Suponha que um indivíduo seja selecionado aleatoriamente da população e defina os eventos A = {tipo A selecionado}, B = {tipo B selecionado} e C = (grupo étnico 3 selecionado).

a. Calcule $P(A)$, $P(C)$, e $P(A \cap C)$.

b. Calcule $P(A \mid C)$ e $P(C \mid A)$ e explique, no contexto, o que significa cada uma dessas probabilidades.

c. Se o indivíduo selecionado não tiver o tipo sangüíneo B, qual é a probabilidade de que ele ou ela seja do grupo étnico 1?

46. Suponha que um indivíduo seja selecionado aleatoriamente da população de todos os homens adultos que vivem nos Estados Unidos. Sejam A o evento em que o indivíduo selecionado tem mais de 1,80m de altura e B o evento em que o indivíduo selecionado é um jogador profissional de basquete. Qual você pensa ser maior: $P(A \mid B)$ ou $P(B \mid A)$? Por quê?

47. Volte ao cenário dos cartões de crédito do Exercício 12 (Seção 2.2), onde A = {Visa}, B = {MasterCard}, $P(A) = 0{,}5$, $P(B) = 0{,}4$ e $P(A \cap B) = 0{,}25$. Calcule e interprete cada uma das probabilidades a seguir (um diagrama de Venn pode ser útil).

a. $P(B \mid A)$ **b.** $P(B' \mid A)$

c. $P(A \mid B)$ **d.** $P(A' \mid B)$

e. Dado que o indivíduo selecionado tem ao menos um cartão de crédito, qual é a probabilidade de ele ter um cartão Visa?

48. Reconsidere a situação de defeitos descrita no Exercício 26 (Seção 2.2).

a. Dado que o sistema tem um defeito do tipo 1, qual é a probabilidade de ele ter um defeito do tipo 2?

b. Dado que o sistema tem um defeito do tipo 1, qual é a probabilidade de ele ter os três tipos de defeitos?

c. Dado que o sistema tem ao menos um tipo de defeito, qual é a probabilidade de ele ter exatamente um tipo de defeito?

d. Dado que o sistema tem os dois primeiros tipos de defeitos, qual é a probabilidade de ele não ter o terceiro tipo de defeito?

49. Se duas lâmpadas forem selecionadas aleatoriamente da caixa do Exercício 38 (Seção 2.3) e pelo menos uma delas for de 75W, qual é a probabilidade de que ambas gerem 75W? Dado que pelo menos uma das lâmpadas não seja de 75W, qual é a probabilidade de ambas as lâmpadas selecionadas terem a mesma potência?

50. Uma loja de departamentos vende camisas esportivas em três tamanhos (pequeno, médio e grande), três padrões (xadrez, estampado e listrado) e dois comprimentos de manga (curta e longa). As tabelas a seguir fornecem as proporções de camisas vendidas nas diversas combinações de categorias.

Manga curta

	Padrão		
Tamanho	**Xadrez**	**Estampado**	**Listrado**
P	0,04	0,02	0,05
M	0,08	0,07	0,12
G	0,03	0,07	0,08

Manga longa

	Padrão		
Tamanho	**Xadrez**	**Estampado**	**Listrado**
P	0,03	0,02	0,03
M	0,10	0,05	0,07
G	0,04	0,02	0,08

a. Qual é a probabilidade de a próxima camisa vendida ser média, de mangas longas e estampada?

b. Qual é a probabilidade de a próxima camisa vendida ser uma camisa média, estampada?

c. Qual é a probabilidade de a próxima camisa vendida ser de mangas curtas? E de mangas longas?

d. Qual é a probabilidade de a próxima camisa vendida ser média? E de ser estampada?

e. Dado que a última camisa vendida foi uma xadrez de mangas curtas, qual é a probabilidade de o tamanho ser médio?

f. Dado que a última camisa vendida foi uma média e xadrez, qual é a probabilidade de as mangas serem curtas? E de serem longas?

51. Uma caixa contém seis bolas vermelhas e três verdes e uma segunda caixa contém sete bolas vermelhas e três verdes. Uma bola é retirada da primeira caixa e colocada na segunda. Então uma bola é retirada da segunda caixa e colocada na primeira.

a. Qual é a probabilidade de uma bola vermelha ser selecionada na primeira caixa e outra bola vermelha na segunda?

b. No fim do processo de seleção, qual é a probabilidade de o número de bolas vermelhas e verdes da primeira e da segunda caixas ser idêntico ao do início?

52. Um sistema consiste em duas bombas idênticas, nº 1 e nº 2. Se uma bomba falha, o sistema continua em operação. Entretanto, devido ao esforço adicional, a bomba restante passa a estar mais suscetível a falhas do que originalmente. Isto é, $r = P(\text{nº 2 falha} \mid \text{nº 1 falha}) > P(\text{nº 2 falha}) = q$. Se pelo menos uma das bombas falhar durante a vida útil do sistema 7% das vezes e ambas as bombas falharem durante aquele período em apenas 1% das vezes, qual é a probabilidade de a bomba nº 1 falhar durante a vida útil do sistema?

53. Certa loja faz reparos em componentes de áudio e vídeo. Represente por A o evento em que o próximo

componente trazido para conserto seja de áudio e por B o evento em que o próximo componente seja um CD-player (de forma que o evento B está contido em A). Suponha que $P(A) = 0{,}6$ e $P(B) = 0{,}05$. Qual é $P(B \mid A)$?

54. No Exercício 13, A_i = {projeto escolhido i}, para $i = 1, 2, 3$. Use as probabilidades fornecidas para calcular as probabilidades a seguir:

a. $P(A_2 \mid A_1)$

b. $P(A_2 \cap A_3 \mid A_1)$

c. $P(A_2 \cup A_3 \mid A_1)$

d. $P(A_1 \cap A_2 \cap A_3 \mid A_1 \cup A_2 \cup A_3)$.

Expresse em palavras a probabilidade calculada.

55. No Exercício 42, seis pessoas (três casais) escolheram assentos aleatoriamente em uma fileira formada por seis assentos.

a. Use a regra da multiplicação para calcular a probabilidade de Jim e Paula se sentarem juntos à esquerda (evento A) e de John e Mary Lou (marido e mulher) se sentarem juntos no meio (evento B).

b. Dado que John e Mary Lou sentam-se juntos no meio, qual é a probabilidade de os outros dois maridos se sentarem junto a suas esposas?

c. Dado que John e Mary Lou sentam-se juntos, qual é a probabilidade de todos os maridos se sentarem junto a suas esposas?

56. Se $P(B \mid A) > P(B)$, mostre que $P(B' \mid A) < P(B')$. (*Sugestão*: Adicione $P(B' \mid A)$ aos dois lados da inequação fornecida e então use o resultado do Exercício 57.)

57. Para quaisquer eventos A e B com $P(B) > 0$, mostre que $P(A \mid B) + P(A' \mid B) = 1$.

58. Demonstre que para quaisquer três eventos A, B, e C com $P(C) > 0$, $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C) - P(A \cap B \mid C)$.

59. Em determinado posto de gasolina, 40% dos clientes usam gasolina comum (A_1), 35% usam gasolina aditivada (A_2) e 25% usam gasolina *premium* (A_3). Dos clientes que usam gasolina comum, apenas 30% enchem o tanque (evento B). Dos clientes que usam gasolina aditivada e *premium*, 60% enchem o tanque, enquanto dentre os que usam *premium*, 50% enchem o tanque.

a. Qual é a probabilidade de o próximo cliente pedir gasolina aditivada e encher o tanque ($A_2 \cap B$)?

b. Qual é a probabilidade de o próximo cliente encher o tanque?

c. Se o próximo cliente encher o tanque, qual é a probabilidade de pedir gasolina comum? E gasolina aditivada? E gasolina *premium*?

60. Setenta por cento das aeronaves leves que desaparecem em vôo em certo país são localizadas posteriormente. Das aeronaves localizadas, 60% possuem localizador de emergência, enquanto 90% das aeronaves não localizadas não possuem esse dispositivo. Suponha que uma aeronave leve tenha desaparecido.

a. Se ela tiver localizador de emergência, qual é a probabilidade de não ser localizada?

b. Se ela não tiver localizador de emergência, qual é a probabilidade de ser localizada?

61. Componentes de determinado tipo são enviados a um fornecedor em lotes de 10. Suponha que 50% desses lotes não contenham componentes defeituosos, 30% contenham um componente defeituoso e 20% contenham dois componentes defeituosos. Dois componentes de um lote são selecionados aleatoriamente e testados. Quais são as probabilidades associadas a 0, 1 e 2 componentes defeituosos estarem no lote em cada uma das condições a seguir?

a. Nenhum componente testado apresenta defeito.

b. Um de dois componentes testados apresenta defeito. (*Sugestão*: Desenhe um diagrama de árvore com três ramos de primeira geração para os três tipos diferentes de lotes.)

62. Uma empresa que fabrica câmeras de vídeo produz um modelo básico e um modelo luxo. No ano passado, 40% das câmeras vendidas foram do modelo básico. Dos clientes que compraram o modelo básico, 30% compraram uma garantia estendida, enquanto 50% dos compradores do modelo luxo a compraram. Se você souber que um cliente selecionado aleatoriamente possui uma garantia estendida, qual a probabilidade de ele ter o modelo básico?

63. Considere os eventos a seguir para clientes que compram um conjunto de pneus em uma determinada loja, considere os eventos

A = {os pneus comprados foram fabricados nos Estados Unidos}

B = {o comprador fez o balanceamento dos pneus imediatamente}

C = {o comprador solicita alinhamento dianteiro e traseiro}

em conjunto com A', B', e C'. Assuma as probabilidades condicionais e não-condicionais a seguir:

$$P(A) = 0{,}75 \quad P(B \mid A) = 0{,}9 \quad P(B \mid A') = 0{,}8$$

$$P(C \mid A \cap B) = 0{,}8 \quad P(C \mid A \cap B') = 0{,}6$$

$$P(C \mid A' \cap B) = 0{,}7 \quad P(C \mid A' \cap B') = 0{,}3$$

a. Construa um diagrama de árvore formado por ramos de primeira, segunda e terceira gerações, identifique cada evento e a probabilidade apropriada próximo a cada ramo.

b. Calcule $P(A \cap B \cap C)$.

c. Calcule $P(B \cap C)$.

d. Calcule $P(C)$.

e. Calcule $P(A \mid B \cap C)$, a probabilidade de uma compra de pneus americanos, dado que são solicitados alinhamento e balanceamento.

64. No Exemplo 2.30, suponha que a taxa de incidência da doença seja de 1 em 25, em vez de 1 em 1000. Qual

será então a probabilidade de um resultado positivo do teste? Dado que o resultado é positivo, qual a probabilidade de o indivíduo ter a doença? Dado um resultado negativo, qual a probabilidade de o indivíduo não ter a doença?

65. Em uma grande universidade, na pesquisa por um livro-texto satisfatório, o departamento de estatística testou um livro diferente durante os últimos três trimestres. No trimestre do outono, 500 alunos usaram o livro do professor Média; durante o inverno, 300 alunos usaram o livro do professor Mediana e durante a primavera, 200 alunos usaram o livro do professor Moda. Uma pesquisa no final de cada trimestre mostrou que 200 estavam satisfeitos com o livro do Média, 150 estavam satisfeitos com o livro do Mediana e 160 estavam satisfeitos com o livro do Moda. Se um aluno que estudou estatística durante um dos trimestres for selecionado aleatoriamente e admitir estar satisfeito com o livro-texto, ele terá mais probabilidade de ter usado o livro do Média, do Mediana ou do Moda? Quem é o autor igualmente menos provável? (*Sugestão*: Desenhe um diagrama de árvore ou use o teorema de Bayes.)

66. Uma amiga que vive em Los Angeles faz viagens freqüentes de consultoria para Washington, D.C. Cinqüenta por cento das vezes ela viaja pela linha aérea nº 1, 30 % das vezes pela nº 2 e os 20 % restantes das vezes pela nº 3. Na linha nº 1, os vôos para Washington atrasam 30% das vezes e para L.A., 10% das vezes. Na linha aérea nº 2, essas porcentagens são 25% e 20%, enquanto na linha aérea nº 3 são 40% e 25%. Se soubermos que em uma viagem ela chega atrasada em exatamente um dos destinos, quais são as probabilidades posteriores de ela ter voado nas linhas nº 1, nº 2 e nº 3? (*Sugestão*: Na extremidade de cada ramo de primeira geração em um diagrama de árvore, desenhe três ramos de segunda geração denominados 0 atrasado, 1 atrasado e 2 atrasado.)

67. No Exercício 59, considere as informações adicionais a seguir sobre a utilização de cartões de crédito:

70% de todos os clientes da gasolina comum que enchem o tanque usam um cartão de crédito.

50% de todos os clientes da gasolina comum que não enchem o tanque usam um cartão de crédito.

60% de todos os clientes da gasolina extra que enchem o tanque usam um cartão de crédito.

50% de todos os clientes da gasolina extra que não enchem o tanque usam um cartão de crédito.

50% de todos os clientes da gasolina *premium* que enchem o tanque usam um cartão de crédito.

40% de todos os clientes da gasolina *premium* que não enchem o tanque usam um cartão de crédito.

Calcule a probabilidade de cada um dos eventos a seguir para o próximo cliente a ser atendido (um diagrama de árvore pode ajudar).

a. {extra, tanque cheio e cartão de crédito}
b. {*premium*, não tanque cheio e cartão de crédito}
c. {*premium* e cartão de crédito}
d. {tanque cheio e cartão de crédito}
e. {cartão de crédito}
f. se o próximo cliente usar cartão de crédito, qual a probabilidade de que seja pedida a gasolina *premium*?

2.5 Independência

A definição de probabilidade condicional nos permite revisar a probabilidade $P(A)$ originalmente atribuída a A quando somos informados posteriormente da ocorrência de outro evento B. A nova probabilidade de A é $P(A \mid B)$. Em nossos exemplos, normalmente ocorria de $P(A \mid B)$ ser diferente da probabilidade não-condicional $P(A)$, indicando que a informação "B ocorreu" resultou em uma alteração da possibilidade de ocorrência de A. Há outras situações em que a possibilidade de ocorrência de A não é afetada pelo conhecimento da ocorrência de B, de forma que $P(A \mid B) = P(A)$. É natural, então, pensar em A e B como eventos independentes, indicando que a ocorrência ou não-ocorrência desses eventos não afeta a possibilidade de ocorrer o outro evento.

DEFINIÇÃO — Dois eventos A e B são **independentes** se $P(A \mid B) = P(A)$ e **dependentes** em caso contrário.

A definição de independência pode parecer "assimétrica" porque não se exige também que $P(B \mid A) = P(B)$. Entretanto, usando a definição de probabilidade condicional e a regra da multiplicação,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A)} \tag{2.7}$$

O lado direito da Equação (2.7) é $P(B)$ se, e somente se, $P(A \mid B) = P(A)$ (independência), de forma que a igualdade da definição implica outra igualdade (e vice-versa). Também é fácil demonstrar que, se A e B forem independentes, os seguintes pares de eventos também o serão: (1) A' e B, (2) A e B', e (3) A' e B'.

Exemplo 2.31

Considere o lançamento de um dado uma vez e defina os eventos $A = \{2, 4, 6\}$, $B = \{1, 2, 3\}$, e $C = \{1, 2, 3, 4\}$. Temos então $P(A) = \frac{1}{2}$, $P(A \mid B) = \frac{1}{3}$ e $P(A \mid C) = \frac{1}{2}$. Isto é, os eventos A e B são dependentes, enquanto os eventos A e C são independentes. Intuitivamente, se esse dado for lançado e formos informados de que o resultado foi 1, 2, 3 ou 4 (C ocorreu), a probabilidade de A ter ocorrido é $\frac{1}{2}$, conforme era inicialmente, já que dois dos quatro resultados relevantes são pares e são igualmente prováveis. ■

Exemplo 2.32

Sejam A e B dois eventos mutuamente exclusivos com $P(A) > 0$. Por exemplo: para um automóvel escolhido aleatoriamente, seja $A = \{$o carro possui quatro cilindros$\}$ e $B = \{$o carro possui seis cilindros$\}$. Como os eventos são mutuamente exclusivos, se B ocorrer, A não pode ter ocorrido, de forma que $P(A \mid B) = 0 \neq P(A)$. A mensagem aqui é: *se dois eventos forem mutuamente exclusivos, não podem ser independentes*. Quando A e B são mutuamente exclusivos, a informação da ocorrência de A indica algo sobre B (não pode ter ocorrido), de forma que a independência é excluída. ■

$P(A \cap B)$ Quando os eventos são independentes

Freqüentemente a natureza de um experimento indica que dois eventos A e B devem ser considerados independentes. Este é o caso, por exemplo, de um fabricante que recebe uma placa de circuito de cada um de dois fornecedores diferentes; cada placa é testada em sua chegada e $A = \{$a primeira tem defeito$\}$ e $B = \{$a segunda tem defeito$\}$. Se $P(A) = 0{,}1$, também deve ser o caso de $P(A \mid B) = 0{,}1$, sabendo que a condição da segunda placa não fornece informações sobre as condições da primeira. Nosso próximo resultado mostra como calcular $P(A \cap B)$ quando os eventos são independentes.

PROPOSIÇÃO A e B são independentes se, e somente se,

$$P(A \cap B) = P(A) \cdot P(B) \tag{2.8}$$

Parafraseando a proposição, A e B são eventos independentes se, e somente se, a probabilidade de ambos ocorrerem ($A \cap B$) for o produto de duas probabilidades individuais. Segue a demonstração:

$$P(A \cap B) = P(A \mid B) \cdot P(B) = P(A) \cdot P(B) \tag{2.9}$$

onde a segunda igualdade da Equação (2.9) é válida se, e somente se, A e B forem independentes. Por causa da equivalência de independência pela Equação (2.8), esta última pode ser usada como definição de independência.

Exemplo 2.33

Sabe-se que 30% das lavadoras de roupa de uma determinada empresa requerem manutenção enquanto estiverem na garantia e somente 10% das secadoras de roupa precisam de manutenção. Se alguém comprar uma lavadora e uma secadora de roupas feitas por essa empresa, qual é a probabilidade de que ambas as máquinas precisem de conserto?

Seja A o evento em que a lavadora precisa de manutenção em garantia e B o evento análogo para a secadora. Então $P(A) = 0{,}30$ e $P(B) = 0{,}10$. Assumindo que as máquinas funcionem de forma independente uma da outra, a probabilidade desejada será

$$P(A \cap B) = P(A) \cdot P(B) = (0{,}30)(0{,}10) = 0{,}03$$

A probabilidade de que nenhuma das máquinas precise de serviço é

$$P(A' \cap B') = P(A') \cdot P(B') = (0{,}70)(0{,}90) = 0{,}63$$ ■

Exemplo 2.34

Cada dia, de segunda a sexta-feira, chega em uma determinada instalação de inspeção um lote de componentes enviado por um fornecedor. Duas vezes por semana, chega um lote de um segundo fornecedor. Oitenta por cento de todos os lotes do fornecedor 1 passam na inspeção, assim como 90% dos lotes do fornecedor 2. Qual é a probabilidade de, em um dia selecionado aleatoriamente, dois lotes passarem na inspeção? Responderemos a essa questão assumindo que, em dias em que os dois lotes são testados, os fatos de o primeiro ou de o segundo lote passarem ou não na inspeção são independentes. A Figura 2.13 exibe as informações relevantes.

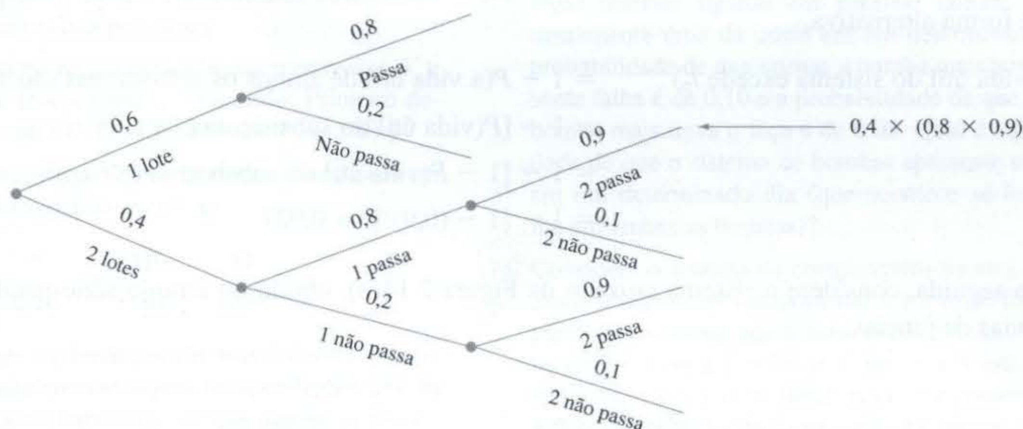


Figura 2.13 Diagrama de árvore do Exemplo 2.34

$$\begin{aligned} P(\text{ambos passam}) &= P(\text{dois recebidos} \cap \text{ambos passam}) \\ &= P(\text{ambos passam} \mid \text{dois recebidos}) \cdot P(\text{dois recebidos}) \\ &= [(0{,}8)(0{,}9)](0{,}4) = 0{,}288 \end{aligned}$$ ■

Independência de mais de dois eventos

A noção de independência de dois eventos pode ser estendida para coleções de mais de dois eventos. Apesar de ser possível estender a definição para dois eventos independentes trabalhando em termos de probabilidades condicionais e não-condicionais, é mais direto e menos trabalhoso seguir a linha da última proposição.

DEFINIÇÃO Os eventos $A_1, \ldots, A_n$ são **mutuamente independentes** se para cada k ($k = 2, 3, \ldots, n$) e cada subconjunto de índices $i_1, i_2, \ldots, i_k$,

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \cdots \cdot P(A_{i_k}).$$

Parafraseando a definição, os eventos serão mutuamente independentes se a probabilidade da interseção de qualquer subconjunto dos n eventos for igual ao produto das probabilidades individuais. Como foi o caso de dois eventos, freqüentemente especificamos a independência de certos eventos no início do problema. A definição pode então ser usada para calcular a probabilidade de uma interseção.

Exemplo 2.35

O artigo "Reliability Evaluation of Solar Photovoltaic Arrays" (*Solar Energy*, 2002, p. 129-141) apresenta diversas configurações de matrizes fotovoltáticas solares formados por células solares de silício cristalino. Considere o sistema ilustrado na Figura 2.14 (a). Há dois subsistemas ligados em paralelo, cada um com três células. Para que o sistema funcione, pelo menos dois subsistemas paralelos devem funcionar. Em cada subsistema, as três células estão ligadas em série, de forma que um subsistema só funciona se todas as suas células estiverem funcionando. Considere a vida útil t_0 e suponha que se queira determinar a probabilidade da vida útil do sistema exceder t_0. Seja A_i o evento vida útil da célula i excede t_0 ($i = 1, 2, ..., 6$). Assumimos que os eventos A_i sejam independentes (se alguma célula específica durar mais de t_0 horas não tem nenhuma relação com a duração das outras células) e que $P(A_i) = 0{,}9$ para cada i, já que as células são idênticas. Então

$$\begin{aligned} P(\text{vida útil do sistema excede } t_0) &= P[(A_1 \cap A_2 \cap A_3) \cup (A_4 \cap A_5 \cap A_6)] \\ &= P(A_1 \cap A_2 \cap A_3) + P(A_4 \cap A_5 \cap A_6) - \\ &\quad P[(A_1 \cap A_2 \cap A_3) \cap (A_4 \cap A_5 \cap A_6)] \\ &= (0{,}9)(0{,}9)(0{,}9) + (0{,}9)(0{,}9)(0{,}9) - (0{,}9)(0{,}9)(0{,}9)(0{,}9)(0{,}9)(0{,}9) = 0{,}927 \end{aligned}$$

De forma alternativa,

$$\begin{aligned} P(\text{vida útil do sistema excede } t_0) &= 1 - P(\text{a vida útil de ambos os subsistemas são} \le t_0) \\ &= 1 - [P(\text{vida útil do subsistema é} \le t_0)]^2 \\ &= 1 - [1 - P(\text{vida útil do subsistema é} > t_0)]^2 \\ &= 1 - [1 - (0{,}9)^3]^2 = 0{,}927 \end{aligned}$$

Em seguida, considere o sistema cruzado da Figura 2.14(b), obtido do arranjo série-paralelo pela ligação das colunas de junção.

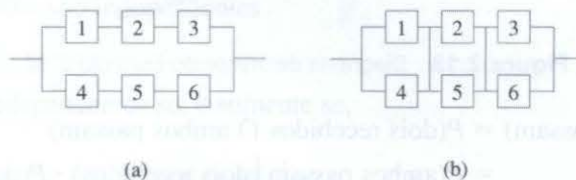


Figura 2.14 Configurações de sistema do Exemplo 2.35: (a) série-paralelo; (b) cruzado

Neste caso o sistema falhará se houver falha de uma coluna, e a vida útil do sistema excederá t_0 apenas se a vida útil de cada coluna também exceder essa configuração,

$$\begin{aligned} P(\text{vida útil do sistema é no mínimo } t_0) &= [P(\text{vida útil da coluna excede } t_0)]^3 \\ &= [1 - P(\text{vida útil da coluna é} \le t_0)]^3 \\ &= [1 - P(\text{ambas as células da coluna têm vida útil} \le t_0)]^3 \\ &= [1 - (1 - 0{,}9)^2]^3 = 0{,}970 \end{aligned}$$

Exercícios | Seção 2.5 (68–87)

68. Reconsidere o cenário de cartões de crédito do Exercício 47 (Seção 2.4) e mostre que A e B são dependentes primeiro usando a definição de independência e depois demonstrando que não se aplica a regra da multiplicação.

69. Uma empresa de exploração de petróleo possui dois projetos ativos, um na Ásia e outro na Europa. Sejam por A o evento em que o projeto da Ásia tem sucesso e B o evento em que o projeto da Europa tem sucesso. Suponha que A e B sejam eventos independentes com $P(A) = 0{,}4$ e $P(B) = 0{,}7$.

a. Se o projeto da Ásia não obtiver sucesso, qual é a probabilidade de o projeto da Europa também não obtê-lo? Explique seu raciocínio.

b. Qual é a probabilidade de pelo menos um dos dois projetos ter sucesso?

c. Dado que pelo menos um dos dois projetos obteve sucesso, qual é a probabilidade de apenas o projeto da Ásia ter sucesso?

70. No Exercício 13, há algum A_i independente de qualquer outro A_j? Responda, usando a propriedade da multiplicação para eventos independentes.

71. Se A e B forem eventos independentes, mostre que A' e B também são independentes. [*Sugestão*: Primeiro defina a relação entre $P(A' \cap B)$, $P(B)$, e $P(A \cap B)$.]

72. Suponha que as proporções de fenótipos sangüíneos em uma população sejam as seguintes:

A	B	AB	O
0,42	0,10	0,04	0,44

Assumindo que os fenótipos de dois indivíduos selecionados aleatoriamente sejam independentes um do outro, qual é a probabilidade de que ambos os fenótipos sejam O? Qual é a probabilidade de que os fenótipos de dois indivíduos selecionados aleatoriamente sejam iguais?

73. Uma das hipóteses básicas da teoria de gráficos de controle (consulte o Capítulo 16) é a que pontos locados sucessivamente sejam independentes uns dos outros. Cada ponto locado indica que o processo de fabricação está operando corretamente ou que há algum tipo de problema. Mesmo quando o processo está funcionando bem, há uma pequena probabilidade de que um determinado ponto indique um problema no processo. Suponha que essa probabilidade seja 0,05. Qual é a probabilidade de que pelo menos um de 10 pontos sucessivos indique um problema quando na verdade o processo opera corretamente? Responda esta pergunta para 25 pontos sucessivos.

74. A probabilidade de um aluno errar a marcação de uma pergunta específica num exame de múltipla escolha é 0,1. Se houver 10 perguntas e elas forem marcadas de forma independente, qual é a probabilidade de não ser cometido nenhum erro? E de ser cometido pelo menos um erro? Se houver n perguntas e a probabilidade de um erro de marcação for p em vez de 0,1, forneça expressões para as duas probabilidades.

75. A junta de uma aeronave requer 25 rebites. A junta terá que ser refeita se qualquer um dos rebites estiver defeituoso. Suponha que os defeitos dos rebites sejam independentes um do outro e tenham a mesma probabilidade.

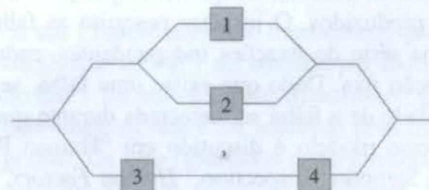
a. Se 20% de todas as juntas tiverem que ser refeitas, qual será a probabilidade de um rebite ter defeito?

b. Quão pequena deve ser a probabilidade de um rebite ter defeito para garantir que apenas 10% das juntas tenham que ser refeitas?

76. Uma caldeira tem cinco válvulas de alívio idênticas. A probabilidade de uma válvula específica ser aberta sob demanda é de 0,95. Assumindo operação independente das válvulas, calcule P(pelo menos uma válvula é aberta) e P(pelo menos uma válvula tem falha ao abrir).

77. Duas bombas ligadas em paralelo falham independentemente uma da outra em um determinado dia. A probabilidade de que apenas a bomba mais antiga apresente falha é de 0,10 e a probabilidade de que apenas a bomba mais nova o faça é de 0,05. Qual é a probabilidade de que o sistema de bombas apresente uma falha em um determinado dia (que acontece se houver falha em ambas as bombas)?

78. Considere o sistema de componentes ligados como na figura a seguir. Os componentes 1 e 2 estão ligados em paralelo, de forma que o subsistema funciona se e, somente se, 1 ou 2 funcionar. Como 3 e 4 estão ligados em série, o subsistema funcionará se, e somente se, 3 e 4 funcionarem. Se os componentes funcionarem independentemente um do outro e P(componente funciona) = 0,9, calcule P(sistema funciona).



79. Refira-se à configuração de sistema série-paralelo apresentada no Exemplo 2.35 e suponha que haja apenas duas células em vez de três em cada subsistema paralelo [na Figura 2.14(a), elimine as células 3 e 6 e renumere as células 4 e 5 como 3 e 4]. Usando $P(A_i) = 0{,}9$, a probabilidade de que a vida útil do sistema exceda t_0 é facilmente vista ser 0,9639. Para que valor 0,9 teria de ser alterado para aumentar a confiabilidade

da vida útil do sistema de 0,9639 para 0,99? (*Sugestão*: Seja $P(A_i) = p$, expresse a confiabilidade do sistema em termos de p e assuma $x = p^2$).

80. Considere o lançamento independente de dois dados, um verde e um vermelho. Sejam A o evento em que o dado vermelho mostra 3 pontos, B o evento em que o dado verde mostra 4 pontos e C o evento em que o número total de pontos mostrados nos dois dados é 7. Esses eventos são independentes aos pares (ou seja, A e B são eventos independentes, A e C também e B e C também)? Os três eventos são mutuamente independentes?

81. Os componentes que chegam em um distribuidor são verificados em busca de defeitos por dois inspetores diferentes (cada componente é verificado por ambos os inspetores). O primeiro inspetor detecta 90% de todos os componentes defeituosos, assim como o segundo inspetor. Pelo menos um inspetor não detecta 20% de todos os componentes defeituosos. Qual é a probabilidade de ocorrência dos itens a seguir?

a. Um componente com defeito ser detectado apenas pelo primeiro inspetor? Por exatamente um dos inspetores?

b. Os três componentes defeituosos de um lote passarem pelos dois inspetores sem detecção de defeito (assumindo que as inspeções de diferentes componentes sejam independentes uma da outra)?

82. Setenta por cento de todos os veículos inspecionados em um determinado posto de inspeção de emissões passam na inspeção. Assumindo que veículos sucessivos passam ou não passam independentemente um do outro, calcule as seguintes probabilidades:

a. P(os três próximos veículos inspecionados passam)

b. P(pelo menos um dos três próximos veículos inspecionados falha)

c. P(exatamente um dos três próximos veículos inspecionados passa)

d. P(no máximo um dos três próximos veículos passa)

e. Dado que ao menos um dos três próximos veículos passa na inspeção, qual é a probabilidade de que todos os três passem (uma probabilidade condicional)?

83. Um inspetor de controle de qualidade pesquisa defeitos em itens produzidos. O inspetor pesquisa as falhas do item numa série de fixações independentes, cada uma com duração fixa. Dado que existe uma falha, seja p a probabilidade de a falha ser detectada durante qualquer fixação (esse modelo é discutido em "Human Performance in Sampling Inspection," *Human Factors*, 1979, p. 99-105).

a. Assumindo que um item tenha uma falha, qual é a probabilidade de ele ser detectado até o final da segunda fixação (depois da detecção de uma falha, as fixações são interrompidas)?

b. Forneça uma expressão da probabilidade de que uma falha será detectada até o final da enésima fixação.

c. Se, quando uma falha não for detectada em três fixações, o item passar, qual será a probabilidade de que um item com falha passe na inspeção?

d. Suponha que 10% de todos os itens contenham uma falha [P(item escolhido aleatoriamente apresenta falha) = 0,1]. Com a suposição da parte (c), qual é a probabilidade de um item com falha passar na inspeção (ele passará automaticamente se não tiver falha, mas também pode passar se tiver)?

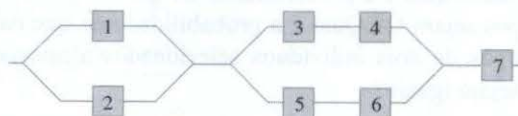
e. Dado que um item passou na inspeção (nenhuma falha em três fixações), qual é a probabilidade de ele possuir uma falha? Calcule para $p = 0,5$.

84. a. Uma serraria recebe um lote de 10.000 tábuas 2 × 4. Suponha que 20% dessas tábuas (2.000) na verdade estejam muito verdes para serem usadas em construção de primeira qualidade. Duas tábuas são selecionadas aleatoriamente, uma após a outra. Sejam A = {a primeira tábua está verde) e B = {a segunda tábua está verde). Calcule $P(A)$, $P(B)$ e $P(A \cap B)$ (um diagrama de árvore é útil). A e B são independentes?

b. Com A e B independentes e $P(A) = P(B) = 0,2$, qual é $P(A \cap B)$? Qual é a diferença entre essa resposta e $P(A \cap B)$ na parte (a)? Para fins de cálculo de $P(A \cap B)$, podemos assumir que A e B da parte (a) sejam independentes para obter unicamente a probabilidade correta?

c. Suponha que o lote consista em 10 tábuas, das quais duas estão verdes. A hipótese de independência agora apresenta aproximadamente a resposta correta para $P(A \cap B)$? Qual é a principal diferença entre a situação descrita aqui e a da parte (a)? Quando você considera que a hipótese de independência seria válida para a obtenção de uma resposta aproximadamente correta para $P(A \cap B)$?

85. Consulte as hipóteses do Exercício 78 e responda à questão proposta para o sistema da figura abaixo. Como a probabilidade mudaria se este fosse um subsistema ligado em paralelo ao subsistema ilustrado na Figura 2.14 (a)?



86. O professor Stan der Deviation pode usar uma de duas rotas em seu caminho de volta do trabalho para casa. Na primeira rota, há quatro cruzamentos com ferrovias. A probabilidade de ele ter de parar em qualquer um dos cruzamentos é 0,1 e os trens operam de forma independente nos quatro cruzamentos. A outra rota é mais longa, mas tem apenas dois cruzamentos, independentes um do outro, com a mesma probabilidade de parada da primeira rota. Um dia, o professor Deviation tem um compromisso em casa com hora marcada. Em qualquer rota ele estima que chegará atrasado se for parado por trens em pelo menos metade dos cruzamentos encontrados.

a. Que rota deve seguir para minimizar a probabilidade de chegar atrasado?

b. Se jogar uma moeda para decidir que rota tomar e chegar atrasado, qual será a probabilidade de ele ter usado a rota com quatro cruzamentos?

87. Suponha que etiquetas idênticas tenham sido colocadas na orelha direita e na esquerda de uma raposa. A raposa é libertada por um período de tempo. Considere os dois eventos C_1 = {etiqueta da orelha esquerda é perdida} e C_2 = {etiqueta da orelha direita é perdida}. Seja $\pi = P(C_1) = P(C_2)$ e assuma C_1 e C_2 como eventos independentes. Deduza uma expressão (envolvendo π) para a probabilidade de exatamente uma etiqueta ser perdida ("Ear Tag Loss in Red Foxes," *J. Wildlife Mgmt.*, 1976, p. 164-167). (*Sugestão*: Desenhe um diagrama de árvore em que os dois ramos iniciais refiram-se ao evento etiqueta da orelha esquerda é perdida).

Exercícios Suplementares (88–111)

88. Uma pequena empresa de fabricação começará a operar no turno da noite. Há 20 maquinistas empregados.

a. Se o turno da noite tiver 3 maquinistas, quantas equipes podem ser formadas?

b. Se os maquinistas forem classificados de 1, 2, ..., 20 em ordem de competência, quantas equipes não teriam o melhor maquinista?

c. Quantas equipes teriam pelo menos um dos 10 melhores maquinistas?

d. Se uma dessas equipes for selecionada aleatoriamente em uma noite específica, qual é a probabilidade de o melhor maquinista não trabalhar naquela noite?

89. Uma empresa usa três linhas de produção para fabricar certo tipo de latas. A tabela fornece as porcentagens de latas fora da conformidade, categorizadas por tipo de não-conformidade, para cada uma das três linhas durante um período de tempo.

	Linha 1	Linha 2	Linha 3
Mancha	15	12	20
Trinca	50	44	40
Problema em *pull-tab*	21	28	24
Defeito na superfície	10	8	15
Outro	4	8	2

Durante esse período, a linha 1 produziu 500 latas fora de conformidade, a linha 2 produziu 400 e a linha 3, 600. Suponha que uma dessas 1500 latas seja selecionada aleatoriamente.

a. Qual é a probabilidade de ela ter sido produzida pela linha 1? E que o motivo da não-conformidade seja uma trinca?

b. Se a lata selecionada tiver vindo da linha 1, qual é a probabilidade de ela ter uma mancha?

c. Dado que a lata selecionada possui um defeito na superfície, qual é a probabilidade de ter vindo da linha 1?

90. Um funcionário do escritório de registros de uma universidade possui 10 formulários em sua mesa esperando processamento. Seis deles são petições de saída e os outros quatro são solicitações de substituição de curso.

a. Se ele selecionar aleatoriamente seis formulários para entregar a um subordinado, qual é a probabilidade de apenas um dos dois tipos de formulários permanecer na mesa?

b. Suponha que ele tenha tempo para processar apenas quatro desses formulários até o fim do dia. Se esses quatro forem selecionados aleatoriamente um a um, qual é a probabilidade de que cada formulário sucessivo seja de um tipo diferente de seu predecessor?

91. Um satélite tem seu lançamento previsto do cabo Canaveral, na Flórida, e o outro da Base Aérea de Vandenberg, na Califórnia. Sejam A o evento do lançamento de Vandenberg ser feito na data e B o evento análogo do cabo Canaveral. Se A e B forem eventos independentes com $P(A) > P(B)$ e $P(A \cup B) = 0{,}626$, $P(A \cap B) = 0{,}144$, determine os valores de $P(A)$ e $P(B)$.

92. Um transmissor está enviando uma mensagem usando código binário, ou seja, uma seqüência de 0s e 1s. Cada *bit* transmitido (0 ou 1) deve passar por três relés para chegar ao receptor. Em cada relé há uma probabilidade de 0,20 de que o *bit* enviado seja feito de forma diferente do *bit* recebido (uma reversão). Assuma que os relés operem de forma independente um do outro.

Transmissor → Relé 1 → Relé 2 → Relé 3 → Receptor

a. Se um 1 for enviado pelo transmissor, qual será a probabilidade de um 1 ser enviado por todos os relés?

b. Se um 1 for enviado pelo transmissor, qual é a probabilidade de um 1 ser recebido pelo receptor? (*Sugestão*: Os oito resultados experimentais podem ser exibidos em um diagrama de árvore com três gerações de ramos, um para cada relé).

c. Suponha que 70% de todos os *bits* enviados do transmissor sejam 1. Se um 1 for recebido pelo receptor, qual é a probabilidade de um 1 ter sido enviado?

93. O indivíduo A tem um círculo de cinco amigos íntimos (B, C, D, E e F). A ouviu um boato de fora do círculo e convidou os cinco amigos para uma festa para transmitir o boato. Para iniciar, A seleciona um dos cinco aleatoriamente e conta o boato para ele. Esse indivíduo

então seleciona um dos quatro restantes e repete o que ouviu. Continuando, uma nova pessoa é selecionada dentre os que não ouviram o boato e assim sucessivamente até que todos saibam.

a. Qual é a probabilidade de o boato ser repetido na ordem B, C, D, E e F?

b. Qual é a probabilidade de F ser a terceira pessoa a saber do boato?

c. Qual é a probabilidade de F ser a última pessoa a ouvir o boato?

94. Refira-se ao Exercício 93. Se em cada etapa a pessoa que "tem" o boato no momento não souber quem já ouviu e selecionar o próximo receptor aleatoriamente dentre as cinco possibilidades, qual é a probabilidade de F ainda não ter ouvido o boato depois ter sido contado 10 vezes na festa?

95. Um engenheiro químico está interessado em determinar se traços de certa impureza estão presentes em um produto. Um experimento tem probabilidade 0,80 de detectar a presença da impureza. A probabilidade de não detectar a impureza se ela não estiver presente é de 0,90. As probabilidades anteriores de a impureza estar presente e estar ausente são de 0,40 e 0,60, respectivamente. Três experimentos separados resultam em apenas duas detecções. Qual é a probabilidade posterior de a impureza estar presente?

96. Cada competidor de um programa de perguntas e respostas é solicitado a especificar uma de seis possíveis categorias que servirão de tema para as perguntas feitas. Suponha que P (competidor escolhe a categoria i) $= \frac{1}{6}$ e os competidores sucessivos escolham suas categorias independentemente um do outro. Se houver três competidores em cada programa e os três escolherem categorias diferentes, qual é a probabilidade de que exatamente um tenha selecionado a categoria 1?

97. Os fixadores usados na fabricação de aviões são ligeiramente tortos para evitar afrouxarem com a vibração. Suponha que 95% de todos os fixadores passem na inspeção inicial. Dos 5% com falhas, 20% possuem defeitos tão sérios que devem ser sucateados. Os fixadores restantes são enviados para retrabalho, onde 40% não podem ser salvos e são descartados. Os outros 60% são corrigidos pelo processo de ondulação e, depois, passam na inspeção.

a. Qual é a probabilidade de um fixador selecionado aleatoriamente passar na inspeção inicialmente ou após o retrabalho?

b. Dado que um fixador tenha passado na inspeção, qual é a probabilidade de ele ter passado na inspeção inicial e não ter precisado de retrabalho?

98. Um por cento de todos os indivíduos de certa população são portadores de determinada doença. Um teste de diagnóstico para essa doença tem taxa de detecção de 90% para portadores e 5% para não-portadores . Suponha que o teste seja aplicado independentemente a duas amostras de sangue do mesmo indivíduo selecionado aleatoriamente.

a. Qual é a probabilidade de ambos os testes terem o mesmo resultado?

b. Se ambos os testes forem positivos, qual será a probabilidade de o indivíduo selecionado ser um portador?

99. Um sistema consiste em dois componentes. A probabilidade de o segundo componente funcionar de forma satisfatória durante a vida útil do projeto é 0,9, a probabilidade de pelo menos um dos dois componentes funcionar é de 0,96 e a de ambos os componentes funcionarem é de 0,75. Dado que o primeiro componente funciona de forma satisfatória por toda a vida útil do projeto, qual é a probabilidade de o segundo também funcionar?

100. Certa empresa envia 40% de sua correspondência expressa noturna pelo serviço expresso E_1. Dessas remessas, 2% chegam após a hora de entrega garantida (represente o evento "entrega atrasada" por L). Se um registro de um envio expresso noturno for selecionado aleatoriamente no arquivo da empresa, qual será a probabilidade de a remessa ter sido enviada via E_1 e estar atrasada?

101. Refira-se ao Exercício 100. Suponha que 50% das remessas expressas sejam enviadas via serviço expresso E_2 e as 10% restantes sejam enviadas por E_3. Dos enviados por E_2, apenas 1% chega com atraso, enquanto 5% das parcelas enviadas por E_3 chegam com atraso.

a. Qual é a probabilidade de uma parcela selecionada aleatoriamente chegar com atraso?

b. Se uma parcela selecionada aleatoriamente chegar no horário, qual será a probabilidade de ela não ter sido enviada por E_1?

102. Uma empresa usa três linhas de montagem diferentes: A_1, A_2 e A_3 para fabricar certo componente. Dos componentes fabricados pela linha A_1, 5% exigem retrabalho para corrigir um defeito, enquanto 8% dos componentes da linha A_2 exigem retrabalho, assim como 10% de A_3. Suponha que 50% de todos os componentes sejam produzidos pela linha A_1, 30% por A_2 e 20% por A_3. Se um componente selecionado aleatoriamente exigir retrabalho, qual será a probabilidade de ele ser proveniente da linha A_1? Da linha A_2? E da linha A_3?

103. Desprezando a possibilidade de aniversário no dia 29 de fevereiro, suponha que um indivíduo selecionado aleatoriamente tenha iguais probabilidades de ter nascido em qualquer um dos outros 365 dias.

a. Se 10 pessoas forem selecionadas aleatoriamente, qual será a probabilidade de todas terem datas de aniversário diferentes? Que ao menos duas tenham a mesma data de aniversário?

b. Com k no lugar de 10 na parte (a), qual é o menor k para o qual há possibilidade de 50% de duas ou mais pessoas terem a mesma data de aniversário?

c. Se 10 pessoas forem selecionadas aleatoriamente, qual será a probabilidade de que pelo menos duas tenham a mesma data de aniversário ou que pelo menos duas tenham os mesmos três últimos dígitos do número de seguro social (ao todo são 9 dígitos)? [*Nota*: O artigo "Methods for Studying Coincidences" (F. Mosteller and P. Diaconis, J. *Amer. Stat. Assoc.*, 1989, p. 853-861) discute problemas desse tipo].

104. Um método usado para distinguir rochas graníticas (*G*) e basálticas (*B*) é examinar uma parte do espectro infravermelho da energia solar refletida na superfície da rocha. Sejam R_1, R_2 e R_3 as intensidades de espectro medidas em três diferentes comprimentos de onda. Normalmente, para granito, $R_1 < R_2 < R_3$, enquanto para basalto $R_3 < R_1 < R_2$. Quando as medidas são feitas de forma remota (usando aeronaves), podem surgir diversas ordenações de R_is, se a rocha for basalto ou granito. Vôos sobre regiões de composição conhecida resultaram nas seguintes informações:

	Granito	**Basalto**
$R_1 < R_2 < R_3$	60%	10%
$R_1 < R_3 < R_2$	25%	20%
$R_3 < R_1 < R_2$	15%	70%

Suponha que, para uma rocha selecionada aleatoriamente em determinada região, *P*(granito) = 0,25 e *P*(basalto) = 0,75.

a. Mostre que $P(\text{granito} \mid R_1 < R_2 < R_3) > P(\text{basalto} \mid R_1 < R_2 < R_3)$. Se as medidas forneceram $R_1 < R_2 < R_3$, você classificaria a rocha como granito ou como basalto?

b. Se as medidas forneceram $R_1 < R_3 < R_2$, como você classificaria a rocha? Responda à mesma questão para $R_3 < R_1 < R_2$.

c. Usando as regras de classificação indicadas nas partes (a) e (b), ao selecionar uma rocha dessa região, qual é a probabilidade de uma classificação errada? [*Sugestão*: *G* pode ser classificado como *B* ou *B* como *C* e *P*(*B*) e *P*(*G*) são conhecidos].

d. Se *P*(granito) = *p* em vez de 0,25, há valores de *p* (diferentes de 1) para os quais uma pessoa sempre classificaria uma rocha como granito?

105. Uma seqüência de vislumbres é fornecida para que um alvo possa ser detectado. Seja G_i = (o alvo é detectado no enésimo vislumbre), com $p_i = P(G_i)$. Suponha que os G_is sejam eventos independentes e escreva uma expressão da probabilidade de que o alvo foi detectado no final do enésimo vislumbre. (*Nota*: Esse modelo é discutido em "Predicting Aircraft Detectability," *Human Factors*, 1979, p. 277-291.)

106. Num jogo de beisebol da Liga Pequena, o lançador do time A lança *strikes* em 50% das vezes e lança *ball* (bolas ruins) 50% das vezes, os arremessos são independentes um do outro e o lançador nunca atinge o rebatedor. Sabendo disso, o treinador do time B instruiu o rebatedor para não rebater nunca. Calcule a probabilidade de

a. o rebatedor ganhar a base no quarto arremesso (4 *balls* consecutivos);

b. o rebatedor ganhar a base no sexto arremesso (de forma que dois dos cinco primeiros devem ter sido *strikes*) usando um argumento de contagem ou construindo um diagrama de árvore;

c. o rebatedor ganhar a base (4 *balls* antes de 3 *strikes*);

d. o primeiro rebatedor marcar sem ninguém ter sido eliminado (assumindo que cada rebatedor use a estratégia de não rebater).

107. Quatro engenheiros, A, B, C e D, tiveram entrevistas marcadas para as 10 da manhã na sexta-feira, 13 de janeiro, na Random Sampling, Inc. O gerente de RH marcou as quatro entrevistas para as salas l, 2, 3 e 4, respectivamente. Entretanto, a secretária do gerente não sabe disso e os manda a quatro salas, de forma totalmente aleatória (o que mais a fazer!). Qual é a probabilidade de:

a. Os quatro acabarem nas salas certas?

b. Nenhum dos quatro ir para a sala certa?

108. Determinada companhia aérea tem vôos às 10 da manhã de Chicago para Nova York, Atlanta e Los Angeles. Seja *A* o evento em que o vôo para Nova York está cheio e defina os eventos *B* e *C* de forma análoga para os outros vôos. Suponha que *P*(*A*) = 0,6, *P*(*B*) = 0,5, *P*(*C*) = 0,4 e que os três eventos sejam independentes. Qual é a probabilidade de:

a. Os três vôos estarem cheios? Ao menos um dos vôos não estar cheio?

b. Apenas o vôo para Nova York estar cheio? Exatamente um dos três vôos estar cheio?

109. Um gerente de pessoal entrevistará quatro candidatos a um cargo. Eles estão classificados como 1, 2, 3 e 4, em ordem de preferência, e serão entrevistados em ordem aleatória. Entretanto, no final de cada entrevista, o gerente saberá apenas como o candidato atual se compara em relação aos candidatos já entrevistados. Por exemplo: a ordem de entrevistas 3, 4, 1, 2 não fornece informações após a primeira entrevista, mostra que o segundo candidato é pior que o primeiro e que o terceiro é melhor que os dois primeiros. Entretanto, a ordem 3, 4, 2, 1 geraria as mesmas informações após cada uma das três primeiras entrevistas. O gerente deseja contratar o melhor candidato, mas deve tomar uma decisão irrevogável após cada entrevista. Considere a estratégia a seguir: rejeitar automaticamente os primeiros *s* candidatos e então contratar o primeiro candidato subseqüente que é o melhor dentre os entrevistados (se não houver tal candidato, contratar o último entrevistado).

Por exemplo, com *s* = 2, a ordem 3, 4, 1, 2 resultaria na contratação do melhor candidato, enquanto a

ordem 3, 1, 2, 4 não o faria. Dos quatro valores possíveis (0, 1, 2 e 3), qual maximiza P (contratação do melhor)? (*Sugestão*: Escreva as 24 ordens de igual probabilidade: $s = 0$ significa que o primeiro candidato é contratado automaticamente).

110. Considere quatro eventos independentes A_1, A_2, A_3 e A_4 e seja $p_i = P(A_i)$ para $i = 1, 2, 3, 4$. Expresse a probabilidade de que pelo menos um desses quatro eventos ocorra em termos dos p_is e faça o mesmo para a probabilidade de que pelo menos dois eventos ocorram.

111. Uma caixa contém quatro pedaços de papel exatamente de mesmas dimensões: (1) ganha o prêmio 1; (2) ganha o prêmio 2, (3) ganha o prêmio 3 e (4) ganha os prêmios 1, 2 e 3. Um pedaço é selecionado aleatoriamente. Seja $A_1 = \{$ganha o prêmio 1$\}$; $A_2 = \{$ganha o prêmio 2$\}$ e $A_3 = \{$ganha o prêmio 3$\}$. Mostre que A_1 e A_2 são independentes, que A_1 e A_2 são independentes, assim como A_2 e A_3 (isso é independência *aos pares*). Mostre, entretanto, que $P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \cdot P(A_2) \cdot P(A_3)$, de forma que esses eventos *não* sejam mutuamente independentes.

Bibliografia

DURRETT, Richard. *The Essentials of Probability.* Duxbury Press, Belmont, CA. 1993. Uma apresentação concisa, em um nível ligeiramente superior ao deste livro.

MOSTELLER, Frederick, ROURKE, Robert e THOMAS, George, *Probability with Statistical Applications* (2ª ed.), Addison-Wesley, Reading, MA, 1970. Uma introdução muito boa à probabilidade, com muitos exemplos divertidos; bom especialmente em regras de contagem e suas aplicações.

OLKIN, Ingram, DERMAN, Cyrus e GLESER, Leon, *Probability Models and Application* (2ª ed.), Macmillan, Nova York, 1994. Uma abrangente introdução à probabilidade, escrita em um nível ligeiramente mais difícil do que o deste livro, mas com muitos exemplos interessantes.

ROSS, Sheldon, *A First Course in Probability* (6ª ed.), Macmillan, Nova York, 2002. De leitura mais difícil e mais sofisticado matematicamente, esse livro contém muitos exemplos e exercícios interessantes.

WINKLER, Robert, *Introduction to Bayesican Inference and Decision,* Holt, Rinehart & Winston. Nova York. 1972. Excelente introdução à probabilidade subjetiva.

3

Variáveis Aleatórias Discretas e Distribuições de Probabilidade

Introdução

Independentemente de um experimento gerar resultados quantitativos ou qualitativos, os métodos de análise estatística requerem que sejam enfocados cértos aspectos numéricos dos dados (como proporção de amostra x/n, média $\bar{x}$ ou desvio padrão s). O conceito de variável aleatória nos permite passar dos resultados do experimento propriamente ditos para uma função numérica dos resultados. Há dois tipos fundamentalmente diferentes de variáveis aleatórias: variáveis aleatórias discretas e variáveis aleatórias contínuas. Neste capítulo examinaremos as propriedades básicas e discutiremos os exemplos mais importantes das variáveis discretas. O Capítulo 4 enfocará as variáveis aleatórias contínuas.

3.1 Variáveis Aleatórias

Em qualquer experimento há diversas características que podem ser observadas ou medidas, mas na maioria dos casos o experimento enfocará um ou alguns aspectos específicos da amostra. Por exemplo: em um estudo de padrões de transporte de uma área metropolitana, cada indivíduo pode ser consultado sobre distância, sobre o número de pessoas que usam o mesmo veículo, mas não sobre o QI, renda, tamanho da família e outras características do tipo. Como alternativa, um pesquisador pode testar uma amostra de componentes e registrar apenas quantos apresentaram falhas dentro de mil horas, em vez de manter o registro das falhas individuais.

Em geral, cada resultado de um experimento é associado a um número, especificando-se uma regra de associação (por exemplo: a quantidade de componentes que apresentam falha em um período de mil horas em uma amostra de 10 componentes ou o peso total de bagagem para uma amostra de 25 passageiros de uma companhia aérea). Tal regra de associação é denominada **variável aleatória**. Variável porque é possível obter diferentes valores numéricos, e aleatória porque o valor observado depende de qual dos resultados possíveis do experimento é obtido (Veja a Figura 3.1).

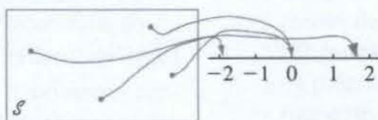


Figura 3.1 Uma variável aleatória

DEFINIÇÃO Para um dado espaço amostral $\mathcal{S}$ de um experimento, uma **variável aleatória (va)** é qualquer regra que associe um valor a cada resultado de $\mathcal{S}$. Em termos matemáticos, uma variável aleatória é uma função cujo domínio é o espaço amostral e o contra-domínio é um conjunto de números reais.

As variáveis aleatórias normalmente são representadas por letras maiúsculas, como X e Y, próximas ao final do alfabeto. Em contraste com a utilização anterior de uma letra maiúscula, como X, para representar uma variável, usaremos letras minúsculas para representar um valor específico da variável aleatória correspondente. A notação $X(s) = x$ significa que x é o valor associado ao resultado s pela va X.

Exemplo 3.1

Quando um estudante tenta acessar um computador em um sistema de compartilhamento de tempo, todas as portas estão ocupadas (F), caso em que o aluno não terá sucesso, ou haverá ao menos uma porta livre (S), caso em que o estudante conseguirá acessar o sistema. Com $\mathcal{S} = \{S, F\}$, defina uma va X como

$$X(S) = 1 \qquad X(F) = 0$$

A va X indica se o estudante pode (1) ou não (0) acessá-lo. ■

No Exemplo 3.1, a va X foi especificada, relacionando-se explicitamente cada elemento de $\mathcal{S}$ e o número associado. Se $\mathcal{S}$ tiver muitos resultados, a tarefa levará muito tempo para ser feita, mas pode ser omitida freqüentemente.

Exemplo 3.2

Considere o experimento em que um número de telefone em um determinado código de área é discado por meio de discador aleatório (dispositivos muito usados por empresas de pesquisa) e defina uma va Y como

$$Y = \begin{cases} 1 & \text{se o número selecionado não estiver na lista telefônica} \\ 0 & \text{se o número selecionado estiver na lista telefônica} \end{cases}$$

Por exemplo: se 5528-2966 estiver na lista telefônica — $Y(55528\text{-}2966) = 0$ — enquanto $Y(77772\text{-}7350) = 1$ nos diz que o número 7772-7350 não está na lista. A descrição desse tipo de evento em palavras é mais econômica do que uma listagem completa, de forma que a usaremos sempre que possível. ■

Nos exemplos 3.1 e 3.2, os únicos valores possíveis da variável aleatória eram 0 e 1. Tais variáveis aleatórias aparecem tão freqüentemente que receberam um nome especial em homenagem ao primeiro indivíduo que as estudou.

DEFINIÇÃO Qualquer variável aleatória cujos únicos valores possíveis são 0 e 1 é denominada **Variável Aleatória de Bernoulli**.

Freqüentemente, precisamos definir e estudar diversas variáveis aleatórias diferentes do mesmo espaço amostral.

Exemplo 3.3

O Exemplo 2.3 descreveu um experimento no qual foi determinado o número de bombas em uso de dois postos de gasolina. Defina as vas *X*, *Y* e *U* por

X = o número total de bombas em uso nos dois postos

Y = a diferença entre o número de bombas em uso no posto 1 e o número em uso no posto 2

U = o máximo de bombas em uso nos dois postos

Se o experimento for realizado e o resultado for $s = (2, 3)$, então $X((2, 3)) = 2 + 3 = 5$, podendo-se dizer que o valor observado de *X* foi $x = 5$. De forma similar, o valor observado de *Y* seria $y = 2 - 3 = -1$ e o valor observado de *U* seria $u = \text{máx.}(2, 3) = 3$. ■

Cada variável aleatória dos exemplos 3.1–3.3 pode assumir apenas um número finito de valores possíveis. Esse não é obrigatoriamente o caso.

Exemplo 3.4

No Exemplo 2.4, consideramos o experimento em que diversas baterias eram examinadas até a obtenção de uma em bom estado (*S*). O espaço amostral foi $\mathcal{S} = \{S, FS, FFS, \ldots\}$. Defina uma va *X* como

X = o número de baterias examinadas antes do fim do experimento

Então $X(S) = 1$, $X(FS) = 2$, $X(FFS) = 3, \ldots, X(FFFFFFS) = 7$ e assim por diante. Qualquer inteiro positivo é um valor possível de *X*, de modo que o conjunto de valores possíveis é infinito. ■

Exemplo 3.5

Suponha que um local (latitude e longitude) nos Estados Unidos continental seja selecionado de forma aleatória. Defina uma va *Y* por

Y = a altura acima do nível do mar do local selecionado

Por exemplo: se o local selecionado foi (39°50′N, 98°35′O), podemos ter *Y*((39°50′N, 98°35′O)) = 1748,26 pés. O maior valor possível de *Y* é 14.494 pés (monte Whitney) e o menor valor possível é −282 pés (Vale da Morte). O conjunto de todos os valores possíveis de *Y* é o conjunto de todos os números no intervalo entre −282 e 14.494, isto é,

$$\{y: y \text{ é um número}, -282 \leq y \leq 14.494\}$$

e há um número infinito de valores nesse intervalo. ■

Dois tipos de variáveis aleatórias

Na Seção 1.2 foi feita uma distinção entre dois tipos de variáveis numéricas, as *discretas* e as *contínuas*. A mesma distinção é feita para as variáveis aleatórias.

DEFINIÇÃO Uma variável aleatória **discreta** é uma variável, cujos valores possíveis constituem um conjunto finito ou podem ser relacionados em uma seqüência infinita na qual haja um primeiro elemento, um segundo e assim por diante. Uma variável aleatória é **contínua** se seu conjunto de valores possíveis consiste em um intervalo completo da reta de números (Reta Real).

Apesar de qualquer intervalo da reta de números conter um número infinito de valores, podemos demonstrar que não há uma forma de criarmos uma lista infinita de todos esses valores, pois há muitos deles.

Exemplo 3.6

Todas as variáveis aleatórias mostradas nos Exemplos 3.1–3.4 são discretas. Como outro exemplo, suponha que selecionemos casais aleatoriamente e façamos um teste sangüíneo em cada pessoa, até encontrarmos um marido e mulher com o mesmo fator Rh. Sendo X = o número de testes sangüíneos a serem realizados, os valores possíveis de X são $D = \{2, 4, 6, 8, \ldots\}$. Como os valores possíveis foram relacionados em seqüência, X é uma va discreta. ■

Para estudar as propriedades básicas das vas discretas, são requeridas apenas as ferramentas da matemática discreta: soma e subtração. O estudo das variáveis contínuas requer a matemática contínua do cálculo: integrais e derivadas.

Exercícios Seção 3.1 (1–10)

1. Uma viga de concreto pode apresentar falha por cisalhamento (C) ou flexão (F). Suponha que três vigas com defeito sejam selecionadas aleatoriamente e o tipo de falha seja determinado para cada uma delas. Seja X = número de vigas entre as três selecionadas que falharam por cisalhamento. Relacione cada resultado no espaço amostral juntamente com o valor de X associado.

2. Dê três exemplos de va de Bernoulli (diferentes dos do livro).

3. Usando o experimento do Exemplo 3.3, defina mais duas variáveis aleatórias e relacione os valores possíveis de cada uma.

4. Seja X = número de dígitos não-nulos de um código CEP selecionado aleatoriamente. Quais são os valores possíveis de X? Forneça três resultados possíveis e seus valores associados X.

5. Se o espaço amostral é um conjunto infinito, implica necessariamente que qualquer va X definida em $\mathcal{S}$ terá um conjunto infinito de valores possíveis? Caso a resposta seja sim, explique o porquê. Caso contrário, dê um exemplo.

6. A partir de certo horário fixo, cada carro que cruza uma interseção é observado para obtenção de sua direção: à esquerda (E), à direita (D) ou segue em frente (F). O experimento termina assim que um carro vira à esquerda. Seja X = o número de carros observados. Quais são os valores possíveis de X? Relacione cinco resultados e seus valores associados X.

7. Para cada variável aleatória definida a seguir, descreva o conjunto de valores possíveis da variável e diga se é discreta.

a. X = número de ovos não quebrados em uma caixa de ovos padrão selecionada aleatoriamente

b. Y = número de estudantes ausentes no primeiro dia de aula em uma lista de chamada de um determinado curso

c. U = número de vezes que um jogador balança o taco de golfe antes de atingir a bola

d. X = comprimento de uma cascavel selecionada aleatoriamente

e. Z = quantia ganha em *royalties* pela venda de uma primeira edição de 10.000 livros-texto

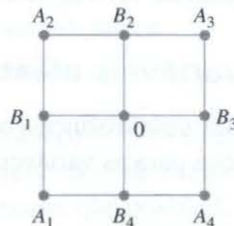
f. Y = pH de uma amostra de solo selecionada aleatoriamente

g. X = tensão (psi) no encordoamento de uma raquete de tênis selecionada aleatoriamente

h. X = número total de lançamentos de moeda exigidos para que três indivíduos obtenham resultados iguais (*cara, cara, cara* ou *coroa, coroa, coroa*)

8. Cada vez que um componente é testado, o teste resulta em sucesso (S) ou falha (F). Suponha que o componente seja testado repetidamente até que ocorra sucesso em três tentativas *consecutivas*. Seja Y o número de tentativas necessárias para atingir esse objetivo. Relacione todos os resultados correspondentes aos cinco menores valores de Y e diga que valor de Y está associado a cada um.

9. Um indivíduo chamado Claudius está localizado no ponto 0 do diagrama a seguir.



Usando um dispositivo de aleatoriedade adequado (como um dado tetraédrico, com quatro lados), Claudius primeiro vai para um dos quatro locais B_1, B_2, B_3, B_4. Uma vez que ele esteja em um desses locais, outro dispositivo aleatório é usado para decidir se ele deve retornar a 0 ou visitar um dos dois outros pontos adjacentes. O processo continua; após cada movimento, outro movimento para um dos (novos) pontos adjacentes é determinado pelo lançamento do dado ou moeda apropriado.

a. Seja X = número de movimentos que Claudius faz antes de retornar a 0. Quais são os valores possíveis de X? X é discreta ou contínua?

b. Se forem permitidos movimentos ao longo da diagonal ligando 0 a A_1, A_2, A_3 e A_4, respectivamente, responda às perguntas da parte (a).

10. O número de bombas em uso em dois postos de gasolina, um com seis bombas e outro com quatro, será determinado. Forneça os valores possíveis de cada uma das variáveis aleatórias a seguir:

a. T = número total de bombas em uso

b. X = diferença entre as quantidades em uso nos postos 1 e 2

c. U = número máximo de bombas em uso em cada posto

d. Z = número de postos com exatamente duas bombas em uso

3.2 Distribuições de Probabilidade para Variáveis Aleatórias Discretas

Quando são atribuídas probabilidades a diversos resultados de $\mathcal{S}$, elas, por sua vez, determinam probabilidades associadas aos valores de qualquer va X em particular. A *distribuição de probabilidade de X* expressa como a probabilidade total 1 é distribuída (alocada a) entre os diversos valores possíveis de X.

Exemplo 3.7

Seis lotes de componentes estão prontos para embarque em um fornecedor. O número de componentes com defeito em cada lote é mostrado a seguir:

Lote	1	2	3	4	5	6
Número de peças com defeito	0	2	0	1	2	0

Um desses lotes será selecionado aleatoriamente para embarque a um cliente específico. Seja X o número de peças com defeito no lote selecionado. Os três valores possíveis de X são 0, 1 e 2. Dos seis eventos igualmente prováveis, três resultam em $X = 0$, um em $X = 1$ e os outros dois em $X = 2$. Sejam $p(0)$ a probabilidade de $X = 0$ e $p(1)$ e $p(2)$ as probabilidades dos outros dois valores possíveis de X. Então

$$p(0) = P(X = 0) = P(\text{lote 1 ou 3 ou 6 é enviado}) = \frac{3}{6} = 0{,}500$$

$$p(1) = P(X = 1) = P(\text{lote 4 é enviado}) = \frac{1}{6} = 0{,}167$$

$$p(2) = P(X = 2) = P(\text{lote 2 ou 5 é enviado}) = \frac{2}{6} = 0{,}333$$

Ou seja, a probabilidade 0,500 é atribuída para o valor 0 de X, a probabilidade 0,167 é alocada para o valor 1 de X e a probabilidade restante, 0,333, é associada ao valor 2 de X. Os valores de X juntamente com suas probabilidades especificam coletivamente a distribuição de probabilidade ou *função de massa da probabilidade de X*. Se o experimento fosse repetido diversas vezes no longo prazo, $X = 0$ ocorreria na metade das vezes, $X = 1$ em um sexto das vezes e $X = 2$ em um terço das vezes. ■

DEFINIÇÃO

A **função distribuição de probabilidade** ou **função de massa de probabilidade** (fmp) de uma va discreta é definida para cada número x por $p(x) = P(X = x) = P(\text{todos os } s \in \mathcal{S}: X(s) = x)$.[1]

[1] $P(X = x)$ lê-se "a probabilidade de a va X assumir o valor x". Por exemplo: $P(X = 2)$ indica a probabilidade de que o valor resultante de X seja 2.

Expondo em palavras, para cada valor possível x da variável aleatória, a fmp especifica a probabilidade de observar aquele valor quando o experimento for realizado. As condições $p(x) \geq 0$ e $\sum_{\text{todos os } x \text{ possíveis}} p(x) = 1$ são obrigatórias para qualquer fmp.

Exemplo 3.8

Suponha que visitemos a livraria de uma universidade durante a primeira semana de aulas, e observemos se a próxima pessoa a comprar um computador comprará um *laptop* ou um *desktop*. Seja

$$X = \begin{cases} 1 & \text{se o cliente comprar um } \textit{laptop} \\ 0 & \text{se o cliente comprar um } \textit{desktop} \end{cases}$$

Se 20% de todos os compradores durante aquela semana selecionarem um *laptop*, a fmp de X será

$$p(0) = P(X = 0) = P(\text{próximo cliente compra um } \textit{desktop}) = 0{,}8$$

$$p(1) = P(X = 1) = P(\text{próximo cliente compra um } \textit{laptop}) = 0{,}2$$

$$p(x) = P(X = x) = 0 \quad \text{para } x \neq 0 \text{ ou } 1$$

Uma descrição equivalente é

$$p(x) = \begin{cases} 0{,}8 & \text{se } x = 0 \\ 0{,}2 & \text{se } x = 1 \\ 0{,}0 & \text{se } x \neq 0 \text{ ou } 1 \end{cases}$$

A Figura 3.2 é uma ilustração dessa fmp, denominada *gráfico de linhas.*

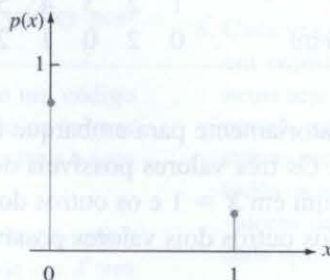


Figura 3.2 O gráfico de linhas da fmp do Exemplo 3.8 ■

Exemplo 3.9

Considere um grupo de cinco doadores de sangue potenciais: A, B, C, D e E. Desses, apenas A e B possuem o tipo O+. Cinco amostras de sangue, uma de cada indivíduo, serão testadas em ordem aleatória até que seja identificado um indivíduo O+. Seja a va Y = número de testes necessários para identificar um indivíduo O+. Então a fmp de Y é

$$p(1) = P(Y = 1) = P(\text{A ou B identificados primeiro}) = \frac{2}{5} = 0{,}4$$

$$p(2) = P(Y = 2) = P(\text{C, D, ou E primeiro e então A ou B})$$

$$= P(\text{C, D, ou E primeiro}) \cdot P(\text{A ou B depois} \mid \text{C, D ou E primeiro}) = \frac{3}{5} \cdot \frac{2}{4} = 0{,}3$$

$$p(3) = P(Y = 3) = P(\text{C, D, ou E primeiro e segundo, então A ou B})$$

$$= \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right) = 0{,}2$$

$p(4) = P(Y = 4) = P(\text{C, D, e E, todos feitos primeiro}) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right) = 0{,}1$

$p(y) = 0 \quad \text{se } y \neq 1, 2, 3, 4$

A fmp pode ser melhor apresentada na forma tabular:

y	1	2	3	4
$p(y)$	0,4	0,3	0,2	0,1

onde qualquer valor y não-relacionado recebe probabilidade zero. Essa fmp também pode ser exibida em um gráfico de linhas (Veja a Figura 3.3).

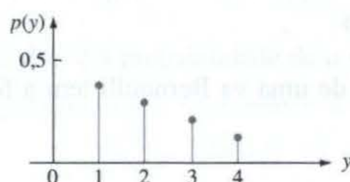


Figura 3.3 O gráfico de linhas da fmp do Exemplo 3.9 ■

O nome "função de massa de probabilidade" é sugerido por um modelo usado em física para um sistema de "pontos de massa". Nesse modelo, as massas são distribuídas nos diversos locais x ao longo de um eixo unidimensional. Nossa fmp descreve como a massa total das probabilidades igual a 1 é distribuída nos diversos pontos ao longo dos eixos dos possíveis valores da variável aleatória (onde e quanto de massa em cada x).

Outra representação ilustrativa útil de uma fmp, denominada histograma de probabilidades, é similar ao histograma descrito no Capítulo 1. Acima de cada y com $p(y) > 0$, construa um retângulo centrado em y. A altura de cada retângulo é proporcional a $p(y)$ e a base é a mesma para todos os retângulos. Quando os valores possíveis estão igualmente espaçados, a base é freqüentemente escolhida como sendo a distância entre os valores sucessivos de y (apesar de poder ser menor). A Figura 3.4 mostra dois histogramas de probabilidades.

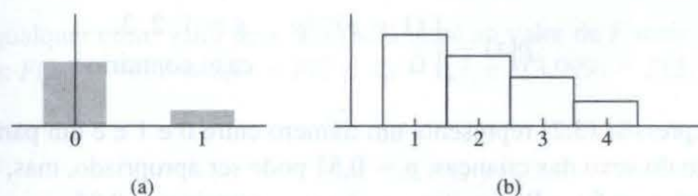


Figura 3.4 Histogramas de probabilidade: (a) Exemplo 3.8; (b) Exemplo 3.9

Parâmetro de uma distribuição de probabilidades

No Exemplo 3.8, tínhamos $p(0) = 0{,}8$ e $p(1) = 0{,}2$ porque 20% de todos os compradores selecionaram computadores portáteis. Em outra livraria, pode ocorrer que $p(0) = 0{,}9$ e $p(1) = 0{,}1$. De forma mais geral, a fmp de qualquer va Bernoulli pode ser expressa na forma $p(1) = \alpha$ e $p(0) = 1 - \alpha$, onde $0 < \alpha < 1$. Como a fmp depende do valor específico de α, normalmente escrevemos $p(x; \alpha)$ em vez de apenas $p(x)$:

$$p(x; \alpha) = \begin{cases} 1 - \alpha & \text{se } x = 0 \\ \alpha & \text{se } x = 1 \\ 0 & \text{caso contrário} \end{cases} \tag{3.1}$$

Cada escolha de α na Expressão (3.1) resulta em uma fmp diferente.

DEFINIÇÃO Suponha que $p(x)$ dependa de uma quantidade que pode ser atribuída a qualquer um de diversos valores possíveis, em que cada valor diferente define uma distribuição de probabilidade diferente. Tal quantidade é denominada **parâmetro** da distribuição. A coleção de todas as distribuições de probabilidade dos diferentes valores do parâmetro é denominada uma **família** de distribuições de probabilidade.

A quantidade α da Expressão (3.1) é um parâmetro. Cada número α entre 0 e 1 determina um membro diferente de uma família para distribuições. Dois desses membros são

$$p(x; 0{,}6) = \begin{cases} 0{,}4 & \text{se } x = 0 \\ 0{,}6 & \text{se } x = 1 \\ 0 & \text{caso contrário} \end{cases} \quad \text{e} \quad p(x; 0{,}5) = \begin{cases} 0{,}5 & \text{se } x = 0 \\ 0{,}5 & \text{se } x = 1 \\ 0 & \text{caso contrário} \end{cases}$$

Cada distribuição de probabilidade de uma va Bernoulli tem a forma da Expressão (3.1), denominada a *família de distribuições de Bernoulli.*

Exemplo 3.10

Iniciando em um horário fixo, observamos o sexo de cada criança nascida em um determinado hospital até que nasça um menino *(B)*. Seja $p = P(B)$, assuma que nascimentos sucessivos sejam independentes e defina a va como X = número de nascimentos observados. Então

$$p(1) = P(X = 1) = P(B) = p$$
$$p(2) = P(X = 2) = P(GB) = P(G) \cdot P(B) = (1 - p)p$$

e

$$p(3) = P(X = 3) = P(GGB) = P(G) \cdot P(G) \cdot P(B) = (1 - p)^2 p$$

Continuando dessa forma, define-se a fórmula geral:

$$p(x) = \begin{cases} (1 - p)^{x-1} p & x = 1, 2, 3, \ldots \\ 0 & \text{caso contrário} \end{cases} \quad (3.2)$$

A quantidade p na Expressão (3.2) representa um número entre 0 e 1 e é um parâmetro da distribuição de probabilidades. No exemplo do sexo das crianças, $p = 0{,}51$ pode ser apropriado, mas, se estivermos buscando a primeira criança com sangue com fator Rh positivo, podemos assumir $p = 0{,}85$. ■

A função distribuição acumulada

Para um valor fixo x, normalmente desejamos computar a probabilidade de o valor observado de X ser no máximo x. Por exemplo: a fmp do Exemplo 3.7 foi

$$p(x) = \begin{cases} 0{,}500 & x = 0 \\ 0{,}167 & x = 1 \\ 0{,}333 & x = 2 \\ 0 & \text{caso contrário} \end{cases}$$

A probabilidade de X ser no máximo 1 é então

$$P(X \leq 1) = p(0) + p(1) = 0{,}500 + 0{,}167 = 0{,}667$$

Nesse exemplo, $X \leq 1{,}5$ se e somente se $X \leq 1$, de forma que $P(X \leq 1{,}5) = P(X \leq 1) = 0{,}667$. De forma similar, $P(X \leq 0) = P(X = 0) = 0{,}5$, e $P(X \leq 0{,}75) = 0{,}5$ também. Como 0 é o menor valor possível de X, $P(X \leq -1{,}7) = 0$, $P(X \leq -0{,}0001) = 0$ e assim por diante. O maior valor possível de X é 2, de forma que $P(X \leq 2) = 1$ e, se x for um número maior que 2, $P(X \leq x) = 1$; isto é, $P(X \leq 5) = 1$, $P(X \leq 10{,}23) = 1$ e assim por diante. Observe que $P(X < 1) = 0{,}5 \neq P(X \leq 1)$, pois a probabilidade de X com valor 1 está incluída na última probabilidade, mas não na anterior. Quando X é uma variável aleatória discreta e x é um valor possível de X, $P(X < x) < P(X \leq x)$.

DEFINIÇÃO

A **função de distribuição acumulada** (FDA) $F(x)$ de uma variável aleatória discreta X com fmp $p(x)$ é definida para cada valor de x por

$$F(x) = P(X \leq x) = \sum_{y:\, y \leq x} p(y)$$

Para qualquer valor x, $F(x)$ é a probabilidade de o valor X observado ser no máximo x.

Exemplo 3.11

A fmp de Y (quantidade de classificações sangüíneas) no Exemplo 3.9 foi

y	1	2	3	4
$p(y)$	0,4	0,3	0,2	0,1

Primeiro determinamos $F(y)$ para cada valor do conjunto $\{1, 2, 3, 4\}$ de valores possíveis:

$$F(1) = P(Y \leq 1) = P(Y = 1) = p(1) = 0{,}4$$
$$F(2) = P(Y \leq 2) = P(Y = 1 \text{ ou } 2) = p(1) + p(2) = 0{,}7$$
$$F(3) = P(Y \leq 3) = P(Y = 1 \text{ ou } 2 \text{ ou } 3) = p(1) + p(2) + p(3) = 0{,}9$$
$$F(4) = P(Y \leq 4) = P(Y = 1 \text{ ou } 2 \text{ ou } 3 \text{ ou } 4) = 1$$

Assim sendo, para qualquer outro valor de y, $F(y)$ será igual ao valor de F mais próximo possível de Y à esquerda de y. Por exemplo: $F(2{,}7) = P(Y \leq 2{,}7) = P(Y \leq 2) = 0{,}7$, e $F(3{,}999) = F(3) = 0{,}9$. A FDA é, portanto

$$F(y) = \begin{cases} 0 & \text{se } y < 1 \\ 0{,}4 & \text{se } 1 \leq y < 2 \\ 0{,}7 & \text{se } 2 \leq y < 3 \\ 0{,}9 & \text{se } 3 \leq y < 4 \\ 1 & \text{se } 4 \leq y \end{cases}$$

O gráfico de $F(y)$ *é mostrado na* Figura 3.5.

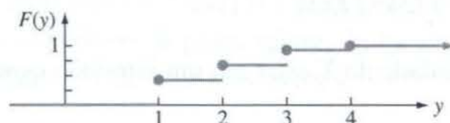


Figura 3.5 O gráfico da FDA do Exemplo 3.11 ■

Para uma va discreta X, o gráfico de $F(x)$ terá um salto em cada valor possível de X e será uma reta entre os valores possíveis. Esse gráfico é denominado **função degrau.**

Exemplo 3.12

No Exemplo 3.10, qualquer inteiro positivo tinha um valor X possível e a fmp era

$$p(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, 3, \ldots \\ 0 & \text{caso contrário} \end{cases}$$

Para qualquer inteiro positivo x,

$$F(x) = \sum_{y \le x} p(y) = \sum_{y=1}^{x} (1-p)^{y-1} p = p \sum_{y=0}^{x-1} (1-p)^y \tag{3.4}$$

Para calcular essa soma, usamos o fato de que a soma parcial de uma série geométrica é

$$\sum_{y=0}^{k} a^y = \frac{1-a^{k+1}}{1-a}$$

Usando essa expressão na Equação (3.4), com $a = 1 - p$ e $k = x - 1$, temos

$$F(x) = p \cdot \frac{1-(1-p)^x}{1-(1-p)} = 1 - (1-p)^x \quad \text{sendo } x \text{ um inteiro positivo}$$

Como F é constante entre inteiros positivos,

$$F(x) = \begin{cases} 0 & x < 1 \\ 1-(1-p)^{[x]} & x \ge 1 \end{cases} \tag{3.5}$$

onde $[x]$ é o maior inteiro $\le x$ (por exemplo, $[2{,}7] = 2$). Portanto, se $p = 0{,}51$, como no exemplo dos nascimentos, a probabilidade de se ter de examinar no máximo cinco nascimentos para se ver o primeiro menino é de $F(5) = 1 - (0{,}49)^5 = 1 - 0{,}0282 = 0{,}9718$, enquanto $F(10) \approx 1{,}0000$. Essa FDA está representada na Figura 3.6.

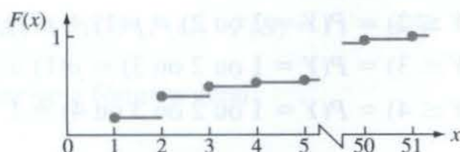


Figura 3.6 Gráfico de $F(x)$ do Exemplo 3.12 ■

Nos exemplos até agora, a FDA origina-se da fmp. Tal processo pode ser revertido para se obter a fmp da FDA sempre que esta última função estiver disponível. Suponha, por exemplo, que X represente o número de componentes com defeito em um lote de seis componentes, de forma que os valores possíveis de X são 0, 1, ..., 6. Então

$$\begin{aligned} p(3) &= P(X = 3) \\ &= [p(0) + p(1) + p(2) + p(3)] - [p(0) + p(1) + p(2)] \\ &= P(X \le 3) - P(X \le 2) \\ &= F(3) - F(2) \end{aligned}$$

De forma mais geral, a probabilidade de X estar em um intervalo especificado é obtida facilmente a partir da FDA. Por exemplo:

$$\begin{aligned} P(2 \le X \le 4) &= p(2) + p(3) + p(4) \\ &= [p(0) + \cdots + p(4)] - [p(0) + p(1)] \\ &= P(X \le 4) - P(X \le 1) \\ &= F(4) - F(1) \end{aligned}$$

Observe que $P(2 \leq X \leq 4) \neq F(4) - F(2)$. O fato ocorre porque o valor 2 de X está incluído em $2 \leq X \leq 4$, de forma que não queremos desprezar essa probabilidade. Entretanto, $P(2 < X \leq 4) = F(4) - F(2)$ porque $X = 2$ não está incluído no intervalo $2 < X \leq 4$.

PROPOSIÇÃO Para quaisquer dois números a e b com $a \leq b$,

$$P(a \leq X \leq b) = F(b) - F(a-)$$

onde "$a-$" representa o maior valor possível de X estritamente menor que a. Em particular, se os únicos valores possíveis forem inteiros e, se a e b forem inteiros, então

$$P(a \leq X \leq b) = P(X = a \text{ ou } a + 1 \text{ ou} \ldots \text{ ou } b)$$
$$= F(b) - F(a - 1)$$

Considerando $a = b$ resulta que $P(X = a) = F(a) - F(a - 1)$, nesse caso.

O motivo de subtrair $F(a-)$ em vez de $F(a)$ é que queremos incluir $P(X = a)$; $F(b) - F(a)$ fornece $P(a < X \leq b)$. Essa proposição será usada extensivamente no cálculo da probabilidade binomial e de Poisson nas Seções 3.4 e 3.6.

Exemplo 3.13

Seja X = o número de dias de licença por doença de um funcionário de uma grande empresa, selecionado aleatoriamente em certo ano. Se o número máximo de dias permitidos por ano for 14, os valores possíveis de X são 0, 1, ..., 14. Com $F(0) = 0{,}58$, $F(1) = 0{,}72$, $F(2) = 0{,}76$, $F(3) = 0{,}81$, $F(4) = 0{,}88$, e $F(5) = 0{,}94$,

$$P(2 \leq X \leq 5) = P(X = 2, 3, 4, \text{ ou } 5) = F(5) - F(1) = 0{,}22$$

e

$$P(X = 3) = F(3) - F(2) = 0{,}05$$

■

Outra visão das funções de massa de probabilidade

Freqüentemente é útil imaginar a fmp como especificação de um modelo matemático para população discreta.

Exemplo 3.14

Considere a escolha aleatória de um estudante entre os 15.000 matriculados no semestre corrente na Mega University. Seja X = número de cursos em que o estudante selecionado está matriculado e suponha que X tenha a fmp mostrada a seguir.

x	1	2	3	4	5	6	7
$p(x)$	0,01	0,03	0,13	0,25	0,39	0,17	0,02

Uma forma de interpretar tal situação é imaginar a população formada por 15.000 indivíduos, cada um com seu próprio valor X. A proporção de cada valor X é dada por $p(x)$. Um ponto de vista alternativo é esquecer dos alunos e pensar na população em si como formada pelos valores X: há alguns 1s na população, 2s e alguns 7s. A população então consiste dos números 1, 2, ..., 7 (portanto, é discreta) e $p(x)$ fornece um modelo da distribuição dos valores da população. ■

Uma vez definido o modelo da população, podemos usá-lo para calcular os valores das características da população (por exemplo, a média μ) e fazer inferências sobre tais características.

Exercícios | Seção 3.2 (11–27)

11. Uma instalação de recondicionamento de automóveis especializada em regulagem de motores sabe que 45% de todas as regulagens são feitas em automóveis de quatro cilindros, 40% em automóveis de seis cilindros e 15% em automóveis de oito cilindros. Seja X = número de cilindros do próximo carro a ser preparado.

a. Qual é a fmp de X?

b. Desenhe um gráfico de linhas e um histograma de probabilidade da fmp da parte (a).

c. Qual é a probabilidade de o próximo carro a ser regulado ter no mínimo seis cilindros? Mais de seis cilindros?

12. As empresas aéreas algumas vezes fazem *overbook* de vôos. Suponha que, para um avião de 50 lugares, tenham sido vendidas 55 passagens. Defina a variável aleatória Y como o número de passageiros com passagens que compareceram ao vôo. A função massa de probabilidade de Y é exibida na tabela a seguir.

y	45	46	47	48	49	50	51	52	53	54	55
$p(y)$	0,05	0,10	0,12	0,14	0,25	0,17	0,06	0,05	0,03	0,02	0,01

a. Qual é a probabilidade de que o vôo acomode todos os passageiros que comparecerem?

b. Qual é a probabilidade de nem todos os passageiros que comparecerem serem acomodados?

c. Se você for a primeira pessoa da lista de espera (o que significa que será o primeiro a subir no avião, se houver assentos disponíveis após todos os passageiros terem sido acomodados), qual será a probabilidade de estar no vôo? Qual é a probabilidade, se você for a terceira pessoa da lista de espera?

13. Uma empresa que fornece computadores pelo correio tem seis linhas telefônicas. Seja X o número de linhas em uso em determinado horário. Suponha que a fmp de X seja conforme a tabela a seguir.

x	0	1	2	3	4	5	6
$p(x)$	0,10	0,15	0,20	0,25	0,20	0,06	0,04

Calcule a probabilidade de cada um dos seguintes eventos.

a. {no máximo três linhas estão em uso}

b. {menos de três linhas estão em uso}

c. {pelo menos três linhas estão em uso}

d. {entre duas e cinco linhas, inclusive, estão em uso}

e. {entre duas e quatro linhas, inclusive, não estão em uso}

f. {pelo menos quatro linhas não estão em uso}

14. Um empreiteiro é solicitado pelo departamento de planejamento de uma cidade a enviar um, dois, três, quatro ou cinco formulários (dependendo da natureza do projeto) quando requer um alvará de construção. Seja Y = o número de formulários requeridos do próximo empreiteiro. Sabe-se que a probabilidade de y formulários serem exigidos é proporcional a y, isto é, $p(y) = ky$ para $y = 1,\ldots,5$.

a. Qual é o valor de k? [*Sugestão*: $\sum_{y=1}^{5} p(y) = 1$.]

b. Qual é a probabilidade de no máximo três formulários serem exigidos?

c. Qual é a probabilidade de serem requeridos entre dois e quatro formulários (inclusive)?

d. Poderia $p(y) = y^2/50$ para $y = 1,\ldots, 5$ ser a fmp de Y?

15. Muitos fabricantes possuem programas de controle de qualidade que incluem a inspeção de defeitos no recebimento dos materiais. Suponha que um fabricante de computadores receba placas em lotes de cinco. Duas placas em cada lote são selecionadas para inspeção. Podemos representar os resultados possíveis do processo pela seleção de pares. Por exemplo: o par (1, 2) representa a seleção das placas 1 e 2 para inspeção.

a. Relacione os 10 resultados diferentes possíveis.

b. Suponha que as placas 1 e 2 sejam as únicas com defeito em um lote de cinco. Duas placas serão escolhidas aleatoriamente. Defina X como o número observado de placas com defeito entre as inspecionadas. Determine a distribuição de probabilidades de X.

c. Seja $F(x)$ a FDA de X. Primeiro determine $F(0) = P(X \leq 0)$, $F(1)$, e $F(2)$, e depois obtenha $F(x)$ para todos os outros x.

16. Algumas partes da Califórnia são particularmente propensas a terremotos. Suponha que em tal área, 30% de todos os moradores tenham seguro contra danos por terremotos. Quatro moradores são selecionados aleatoriamente. Seja X o número dos que possuem seguro contra danos por terremotos, entre os quatro.

a. Determine a distribuição de probabilidades de X. [*Sugestão*: seja S o morador que tem seguro e N o que não o tem. Uma possibilidade será *SNSS*, com probabilidade (0,3)(0,7)(0,3)(0,3) e valor associado $X = 3$. Há outros 15 resultados].

b. Desenhe o histograma de probabilidade correspondente.

c. Qual é o valor mais provável de X?

d. Qual é a probabilidade de que ao menos dois dos quatro moradores selecionados tenham seguro?

17. A voltagem de uma pilha nova pode ser aceitável (A) ou inaceitável (I). Uma lanterna específica exige duas pilhas, que serão selecionadas e testadas independentemente até que sejam encontradas duas aceitáveis. Suponha que 90% de todas as pilhas tenham voltagem aceitável. Seja Y o número de pilhas que devem ser testadas.

a. Qual é o valor de $p(2)$, isto é, $P(Y = 2)$?

b. Qual é o valor de $p(3)$? (*Sugestão:* Há dois resultados diferentes que dão $Y = 3$.)

c. Para ter $Y = 5$, o que deve ser verdadeiro para a quinta pilha selecionada? Relacione os quatro resultados possíveis para os quais $Y = 5$ e depois determine $p(5)$.

d. Use o padrão de suas respostas para as partes (a)–(c) para obter a fórmula geral de $p(y)$.

18. Dois dados de seis lados são lançados independentemente. Seja M = o máximo dos dois lançamentos (então $M(1, 5) = 5$, $M(3, 3) = 3$ etc.).

a. Qual é a fmp de M? [*Sugestão*: Determine primeiro $p(1)$, depois $p(2)$ e assim por diante].

b. Determine a FDA de M e desenhe o gráfico.

19. No Exemplo 3.9, suponha que haja apenas quatro doadores potenciais de sangue dos quais apenas um tem sangue tipo O+. Calcule a fmp de Y.

20. Uma biblioteca assina duas revistas semanais diferentes, que supostamente chegam na correspondência de quarta-feira. Na verdade, elas podem chegar na quarta, quinta, sexta ou sábado. Suponha que as duas cheguem independentemente uma da outra e para cada uma $P(\text{qua.}) = 0{,}3$, $P(\text{qui.}) = 0{,}4$, $P(\text{sex.}) = 0{,}2$ e $P(\text{sáb.}) = 0{,}1$. Seja Y= número de dias, após a quarta-feira, que as revistas levam para chegar (os valores possíveis de Y são 0, 1, 2 ou 3). Calcule a fmp de Y. [*Sugestão*: Há 16 resultados possíveis; $Y(qua., qua.) = 0$, $Y(qua., qui.) = 2$ e assim por diante].

21. Refira-se ao Exercício 13, calcule e desenhe o gráfico da FDA $F(x)$. Utilize-a para calcular a probabilidade dos eventos descritos nas partes (a)–(d) do problema.

22. Uma organização de consumidores que avalia automóveis novos relata costumeiramente o número de defeitos graves em cada carro examinado. Seja X o número de defeitos graves em um carro de determinado tipo selecionado aleatoriamente. A FDA de X é como abaixo:

$$F(x) = \begin{cases} 0 & x < 0 \\ 0{,}06 & 0 \le x < 1 \\ 0{,}19 & 1 \le x < 2 \\ 0{,}39 & 2 \le x < 3 \\ 0{,}67 & 3 \le x < 4 \\ 0{,}92 & 4 \le x < 5 \\ 0{,}97 & 5 \le x < 6 \\ 1 & 6 < x \end{cases}$$

Calcule as seguintes probabilidades diretamente pela FDA:

a. $p(2)$, isto é, $P(X = 2)$

b. $P(X > 3)$

c. $P(2 \le X \le 5)$

d. $P(2 < X < 5)$

23. Uma empresa de seguros oferece aos seus segurados diferentes opções de pagamento *premium*. Para um segurado selecionado aleatoriamente, seja X = número de meses entre pagamentos sucessivos. A FDA de X é como segue:

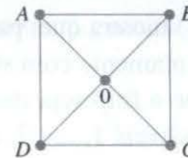
$$F(x) = \begin{cases} 0 & x < 1 \\ 0{,}30 & 1 \le x < 3 \\ 0{,}40 & 3 \le x < 4 \\ 0{,}45 & 4 \le x < 6 \\ 0{,}60 & 6 \le x < 12 \\ 1 & 12 \le x \end{cases}$$

a. Qual é a fmp de X?

b. Usando apenas a FDA, calcule $P(3 \le X \le 6)$ e $P(4 \le X)$.

24. No Exemplo 3.10, seja Y = o número de meninas nascidas antes do final do experimento. Com $p = P(B)$ e $1 - p = P(G)$, qual é a fmp de Y? (*Sugestão*: Primeiro relacione os valores possíveis de Y, começando pelo menor, e continue até obter uma fórmula geral).

25. Alvie Singer vive no ponto 0 do diagrama abaixo e tem quatro amigos que vivem em A, B, C e D. Um dia, Alvie decide visitá-los e lança uma moeda duas vezes para decidir quais dos quatros visitará. Depois de chegar na casa de um amigo, ele pode retornar para casa ou continuar a visita a uma das duas casas adjacentes (como 0, A ou C quando em B), tendo cada uma das três possibilidades com probabilidade $\frac{1}{3}$. Dessa forma, ele continua a visitar amigos até voltar para casa.



a. Seja X = o número de vezes que Alvie visita um amigo. Determine a fmp em relação a X.

b. Seja Y = o número de segmentos de reta que Alvie percorre (incluindo os que chegam e saem de 0). Qual é a fmp de Y?

c. Suponha que amigas vivam em A e C e amigos em B e D. Se Z = o número de visitas a amigas, qual é a fmp de Z?

26. Depois que todos os estudantes saírem da sala de aula, um professor de estatística observa que quatro cópias do livro foram deixadas sob as mesas. No começo da próxima aula, o professor distribui os quatro livros de forma completamente aleatória para cada um de quatro alunos (1, 2, 3 e 4) que dizem ter esquecido os livros. Um resultado possível é que 1 receba o livro de 2, 2 receba o livro de 4, 3 receba o seu livro e 4 receba o livro de 1. Esse resultado pode ser abreviado como (2, 4, 3, 1).

a. Relacione os outros 23 resultados possíveis.

b. Seja X o número de estudantes que recebem o próprio livro. Determine a fmp de X.

27. Demonstre que a FDA $F(x)$ é uma função não-decrescente; isto é, $x_1 < x_2$ implica $F(x_1) \le F(x_2)$. Em que condições $F(x_1) = F(x_2)$?

3.3 Valores Esperados de Variáveis Aleatórias Discretas

No Exemplo 3.14, consideramos uma universidade com 15.000 alunos e consideramos X = o número de cursos em que um aluno selecionado aleatoriamente estava matriculado. A fmp de X é mostrada a seguir. Como $p(1) = 0{,}01$, sabemos que $(0{,}01) \cdot (15.000) = 150$ alunos estão matriculados em um curso e assim sucessivamente para os outros x valores.

x	1	2	3	4	5	6	7
$p(x)$	0,01	0,03	0,13	0,25	0,39	0,17	0,02
Número de alunos matriculados	150	450	1950	3750	5850	2550	300

(3.6)

Para calcular o número médio de cursos por aluno ou o valor médio de X na população, devemos calcular o número total de cursos e dividir pelo número total de alunos. Como cada um dos 150 alunos está matriculado em um curso, esses 150 contribuem com 150 cursos no total. De forma similar, 450 alunos contribuem com 2(450) cursos e assim por diante. O valor médio da população X é

$$\frac{1(150) + 2(450) + 3(1950) + \cdots + 7(300)}{15.000} = 4{,}57 \tag{3.7}$$

Como $150/15.000 = 0{,}01 = p(1)$, $450/15.000 = 0{,}03 = p(2)$, e assim por diante, uma expressão alternativa para (3,7) é

$$1 \cdot p(1) + 2 \cdot p(2) + \cdots + 7 \cdot p(7) \tag{3.8}$$

A Expressão (3.8) mostra que, para calcular o valor médio da população de X, precisamos apenas dos valores possíveis de X juntamente com suas probabilidades (proporções). Em particular, o tamanho da população é irrelevante, desde que a fmp seja dada por (3,6). A média ou valor médio de X é, portanto, uma média *ponderada* dos valores possíveis 1, ... , 7, em que os pesos são as probabilidades dos valores.

Valor esperado de X

DEFINIÇÃO Seja X como uma va discreta com conjunto de valores possíveis D e fmp $p(x)$. O **valor esperado** ou **valor médio** de X, denotado por $E(X)$ ou μ_X, é

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Quando está claro a que X o valor esperado se refere, μ é usado no lugar de μ_X.

Exemplo 3.15

Para a fmp em (3.6),

$$\begin{aligned} \mu &= 1 \cdot p(1) + 2 \cdot p(2) + \cdots + 7 \cdot p(7) \\ &= (1)(0{,}01) + 2(0{,}03) + \cdots + (7)(0{,}02) \\ &= 0{,}01 + 0{,}06 + 0{,}39 + 1{,}00 + 1{,}95 + 1{,}02 + 0{,}14 = 4{,}57 \end{aligned}$$

Se imaginarmos que a população consiste dos X valores 1, 2, ... , 7, então $\mu = 4{,}57$ é a média da população. Na seqüência, freqüentemente nos referiremos a μ como a *média da população* em vez de média de X na população. ■

No Exemplo 3.15, o valor esperado de μ era 4,57, que não é um valor possível de X. A palavra *esperado* deve ser interpretada com cautela porque uma pessoa não espera ver um valor de $X = 4,57$ quando um único estudante é selecionado.

Exemplo 3.16

Após cada nascimento, os bebês são classificados de acordo com uma escala denominada Apgar. As classificações possíveis são 0, 1,..., 10, com a classificação do bebê determinada por cor, tônus muscular, esforço respiratório, batimentos cardíacos e irritabilidade reflexas (a melhor pontuação possível é 10). Seja X o escore Apgar de uma criança selecionada aleatoriamente em um determinado hospital no próximo ano e suponha que a fmp de X seja

x	0	1	2	3	4	5	6	7	8	9	10
$p(x)$	0,002	0,001	0,002	0,005	0,02	0,04	0,18	0,37	0,25	0,12	0,01

O valor médio de X será

$$\begin{aligned} E(X) = \mu &= 0(0{,}002) + 1(0{,}001) + 2(0{,}002) \\ &+ \cdots + 8(0{,}25) + 9(0{,}12) + 10(0{,}01) \\ &= 7{,}15 \end{aligned}$$

Novamente, μ não é um valor possível da variável X. Além disso, como a variável se refere a uma futura criança, não há uma população existente, concreta, a que se refira. Em vez disso, imaginamos a fmp como um modelo para uma população conceitual consistindo dos valores 0, 1, 2,..., 10. O valor médio dessa população conceitual é $\mu = 7,15$. ■

Exemplo 3.17

Seja $X = 1$ se um componente selecionado aleatoriamente precisar de reparo em garantia e $= 0$ caso contrário. Então X será uma va Bernoulli com fmp

$$p(x) = \begin{cases} 1 - p & x = 0 \\ p & x = 1 \\ 0 & x \neq 0, 1 \end{cases}$$

pela qual $E(X) = 0 \cdot p(0) + 1 \cdot p(1) = 0(1 - p) + 1(p) = p$. Isto é, o valor esperado de X é apenas a probabilidade de X ser igual ao valor 1. Se conceitualizarmos uma população formada por zeros na proporção $1 - p$ e de 1s na proporção p, então a média da população é $\mu = p$. ■

Exemplo 3.18

A forma geral da fmp de X = número de crianças nascidas até o nascimento do primeiro menino (inclusive) é

$$p(x) = \begin{cases} p(1 - p)^{x-1} & x = 1, 2, 3, \ldots \\ 0 & \text{caso contrário} \end{cases}$$

Pela definição,

$$E(X) = \sum_D x \cdot p(x) = \sum_{x=1}^{\infty} xp(1 - p)^{x-1} = p \sum_{x=1}^{\infty} \left[-\frac{d}{dp}(1 - p)^x \right] \tag{3.9}$$

Se trocarmos a ordem de cálculo da derivada e da soma, a soma será a de uma série geométrica. Depois de calcular a soma, calcula-se a derivada e o resultado final é $E(X) = 1/p$. Se p estiver próximo de 1, espera-se ver um menino em breve, enquanto, se p estiver próximo de 0, esperam-se muitos nascimentos antes do primeiro menino. Para $p = 0,5$, $E(X) = 2$. ■

Há outra interpretação de μ usada freqüentemente. Considere a fmp

$$p(x) = \begin{cases} (0{,}5) \cdot (0{,}5)^{x-1} & \text{se } x = 1, 2, 3, \ldots \\ 0 & \text{caso contrário} \end{cases}$$

Essa é a fmp de X = número de lançamentos necessários de uma moeda para obter a primeira "cara" (um caso especial do Exemplo 3.18). Suponha que observemos um valor x dessa fmp (lançar uma moeda até aparecer uma "cara"), e então observemos outro valor independentemente (continue lançando), e depois outro e assim sucessivamente. Se depois de observarmos um grande número de valores x calcularmos a média deles, a média amostral resultante será muito próxima de $\mu = 2$. Isto é, μ pode ser interpretado como o valor observado da média de longo prazo de X quando o experimento é executado repetidamente.

Exemplo 3.19

Seja X o número de entrevistas pelas quais um estudante passa antes de conseguir um emprego, com fmp

$$p(x) = \begin{cases} k/x^2 & x = 1, 2, 3, \ldots \\ 0 & \text{caso contrário} \end{cases}$$

onde k é escolhido de forma que $\sum_{x=1}^{\infty} (k/x^2) = 1$. (Nos cursos de matemática sobre séries infinitas, é demonstrado que $\sum_{x=1}^{\infty} (1/x^2) < \infty$, que implica que k existe, mas seu valor exato não nos preocupa). O valor esperado de X é

$$\mu = E(X) = \sum_{x=1}^{\infty} x \cdot \frac{k}{x^2} = k \sum_{x=1}^{\infty} \frac{1}{x} \tag{3.10}$$

A soma à direita da Equação (3.10) é a famosa série harmônica da matemática e pode ser demonstrado que é igual a ∞. $E(X)$ não é finito, neste caso, porque $p(x)$ não decresce tão rapidamente quanto x aumenta. Os estatísticos dizem que a distribuição de probabilidade de X possui "uma cauda longa". Se for escolhida uma seqüência de valores de X usando essa distribuição, a média amostral não terá um valor finito, mas tenderá a crescer sem limite.

Os estatísticos usam a expressão "caudas longas" em relação a quaisquer distribuições que tenham valor de probabilidade grande distante de μ, (portanto caudas longas não requerem $\mu = \infty$). Caudas longas dificultam as inferências sobre μ. ■

Valor esperado de uma função

Freqüentemente estamos interessados no valor esperado de alguma função $h(X)$ em vez de X em si mesmo.

Exemplo 3.20

Suponha que uma livraria compre 10 cópias de um livro a US$ 6,00 cada para vendê-las a US$ 12,00, sabendo que ao fim de um período de 3 meses os livros não vendidos podem ser devolvidos por US$ 2,00. Se X = ao número de cópias vendidas, então a receita líquida será $= h(X) = 12X + 2(10 - X) - 60 = 10X - 40$. ■

Uma forma fácil de calcular o valor esperado de $h(X)$ é sugerida pelo exemplo a seguir.

Exemplo 3.21

Seja X = o número de cilindros do motor do próximo carro a ser regulado em certa oficina. O custo de uma regulagem é relacionado a X por $h(X) = 20 + 3X + 0{,}5X^2$. Como X é uma variável aleatória, $h(X)$ também o é. Denote essa última va por Y. As fmps de X e Y são:

x	4	6	8
$p(x)$	0,5	0,3	0,2

y	40	56	76
$p(y)$	0,5	0,3	0,2

Denotando os valores possíveis de Y por D^*, temos

$$E(Y) = E[h(X)] = \sum_{D^*} y \cdot p(y) \tag{3.11}$$
$$= (40)(0,5) + (56)(0,3) + (76)(0,2)$$
$$= h(4) \cdot (0,5) + h(6) \cdot (0,3) + h(8) \cdot (0,2)$$
$$= \sum_{D} h(x) \cdot p(x)$$

De acordo com a Equação (3.11), não é necessário determinar a fmp de Y para obter $E(Y)$. O valor esperado é a média ponderada dos valores possíveis de $h(x)$ (em vez de x). ■

PROPOSIÇÃO

Se a va X tiver um conjunto de valores possíveis D e fmp $p(x)$, o valor esperado de qualquer função $h(X)$, expresso por $E[h(X)]$ ou $\mu_{h(X)}$, é calculado por

$$E[h(X)] = \sum_{D} h(x) \cdot p(x)$$

De acordo com a proposição, $E[h(X)]$ é calculada da mesma forma que $E(X)$ em si, exceto que $h(x)$ é usado no lugar de x.

Exemplo 3.22

Uma loja de computadores comprou três computadores de certo tipo a US$ 500 cada. Eles serão vendidos a US$ 1.000 cada. O fabricante concordou em aceitar a devolução dos computadores não vendidos, após um período especificado, por US$ 200 cada. Seja X o número de computadores vendidos e suponha que $p(0) = 0,1$, $p(1) = 0,2$, $p(2) = 0,3$, e $p(3) = 0,4$. Definindo como $h(X)$ o lucro associado à venda de X unidades, as informações fornecidas implicam que $h(X)$ = receita − custo $= 1000X + 200(3 - X) - 1500 = 800X - 900$. O lucro esperado será então

$$E[h(X)] = h(0) \cdot p(0) + h(1) \cdot p(1) + h(2) \cdot p(2) + h(3) \cdot p(3)$$
$$= (-900)(0,1) + (-100)(0,2) + (700)(0,3) + (1500)(0,4)$$
$$= \$\ 700$$
■

Regras do valor Esperado

A função $h(X)$ freqüentemente é uma função linear $aX + b$. Nesse caso, $E[h(X)]$ é facilmente calculado pela $E(X)$.

PROPOSIÇÃO

$$E(aX + b) = a \cdot E(X) + b$$

(Ou, usando notação alternativa, $\mu_{aX+b} = a \cdot \mu_X + b$.)

Em palavras, o valor esperado de uma função linear é igual à função linear calculada com o valor esperado $E(X)$. Como $h(X)$ no Exemplo 3.22 é linear e $E(X) = 2$, $E[h(X)] = 800(2) - 900 =$ US$ 700, como antes.

Demonstração

$$E(aX + b) = \sum_{D} (ax + b) \cdot p(x) = a\sum_{D} x \cdot p(x) + b\sum_{D} p(x)$$
$$= aE(X) + b$$
■

Dois casos especiais da proposição fornecem duas regras importantes do valor esperado.

1. Para qualquer constante a, $E(aX) = a \cdot E(X)$ (assumir $b = 0$). (3.12)
2. Para qualquer constante b, $E(X + b) = E(X) + b$ (assumir $a = 1$).

A multiplicação de X por uma constante a altera a unidade de medida (de dólares para centavos, onde $a = 100$, de polegadas para centímetros, onde $a = 2{,}54$ etc.). A regra 1 diz que o valor esperado nas novas unidades é igual ao valor das antigas, multiplicado pelo fator de conversão a. De forma similar, se uma constante b for adicionada a cada valor possível de X, então o valor esperado será deslocado do valor da constante.

Variância de X

O valor esperado de X descreve onde a distribuição de probabilidades está centrada. Usando a analogia física de colocação de pontos de massa $p(x)$ no valor x de um eixo unidimensional, se o eixo estiver apoiado em um fulcro posicionado em μ, não há tendência de inclinação do eixo. O fato é ilustrado para duas distribuições diferentes na Figura 3.7.

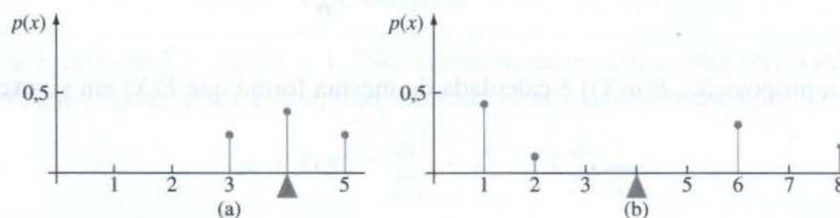


Figura 3.7 Duas distribuições de probabilidade diferentes com $\mu = 4$

Apesar de ambas as distribuições ilustradas na Figura 3.7 terem o mesmo centro μ, a distribuição da Figura 3.7(b) tem maior dispersão ou variabilidade do que a da Figura 3.7(a). Usaremos a variância de X para avaliar o valor da variabilidade de (a distribuição de) X, do mesmo modo que s^2 foi usado no Capítulo 1 para medir a variabilidade de uma amostra.

DEFINIÇÃO

Seja X com fmp $p(x)$ e o valor esperado μ. A **variância** de X, denotada por $V(X)$ ou σ_X^2, ou apenas σ^2, é

$$V(X) = \sum_D (x - \mu)^2 \cdot p(x) = E[(X - \mu)^2]$$

O desvio padrão (DP) de X é

$$\sigma_X = \sqrt{\sigma_X^2}$$

A quantidade $h(X) = (X - \mu)^2$ é o quadrado do desvio de X em relação à sua média e σ^2 é o quadrado do desvio esperado. Se a maior parte da distribuição de probabilidade estiver próxima de μ, σ^2 será relativamente pequena. Entretanto, se houver valores de x distantes de μ com $p(x)$ grande, σ^2 será bem grande.

Exemplo 3.23

Se X é o número de cilindros do próximo carro a ser regulado em uma oficina, com a fmp do Exemplo 3.21 [$p(4) = 0{,}5$, $p(6) = 0{,}3$, $p(8) = 0{,}2$, de onde $\mu = 5.4$], então

$$V(X) = \sigma^2 = \sum_{x=4}^{8} (x - 5{,}4)^2 \cdot p(x)$$
$$= (4 - 5{,}4)^2(0{,}5) + (6 - 5{,}4)^2(0{,}3) + (8 - 5{,}4)^2(0{,}2) = 2{,}44$$

O desvio padrão de X é $\sigma = \sqrt{2{,}44} = 1{,}562$. ■

Quando a fmp $p(x)$ especifica um modelo matemático para a distribuição dos valores da população, σ^2 e σ medem a dispersão dos valores da população: σ^2 é a variância da população e σ é o desvio padrão da população.

Fórmula Alternativa de σ^2

A quantidade de operações aritméticas necessárias para calcular σ^2 pode ser reduzida usando-se uma fórmula alternativa de cálculo.

PROPOSIÇÃO $$V(X) = \sigma^2 = \left[\sum_D x^2 \cdot p(x)\right] - \mu^2 = E(X^2) - [E(X)]^2$$

Ao usar essa fórmula, $E(X^2)$ é calculado primeiro sem necessidade de subtração e então $E(X)$ é calculado, elevado ao quadrado e subtraído (uma vez) de $E(X^2)$.

Exemplo 3.24

A fmp do número de cilindros X do próximo carro a ser regulado em uma oficina foi fornecida no Exemplo 3.23 como $p(4) = 0{,}5$, $p(6) = 0{,}3$, e $p(8) = 0{,}2$, de onde $\mu = 5{,}4$ e

$$E(X^2) = (4^2)(0{,}5) + (6^2)(0{,}3) + (8^2)(0{,}2) = 31{,}6$$

Portanto $\sigma^2 = 31{,}6 - (5{,}4)^2 = 2{,}44$, conforme o Exemplo 3.23. ■

Demonstração da Fórmula Alternativa

Desenvolva $(x - \mu)^2$ da definição de σ^2 para obter $x^2 - 2\mu x + \mu^2$, e depois calcule a somatória ($\sum$) de cada um dos três termos:

$$\sigma^2 = \sum_D x^2 \cdot p(x) - 2\mu \cdot \sum_D x \cdot p(x) + \mu^2 \sum_D p(x)$$

$$= E(X^2) - 2\mu \cdot \mu + \mu^2 = E(X^2) - \mu^2$$ ■

Regras da Variância

A variância de $h(X)$ é o valor esperado do quadrado da diferença entre $h(X)$ e seu valor esperado:

$$V[h(X)] = \sigma^2_{h(X)} = \sum_D \{h(x) - E[h(X)]\}^2 \cdot p(x) \tag{3.13}$$

Quando $h(X)$ for uma função linear, $V[h(X)]$ é facilmente relacionada a $V(X)$.

PROPOSIÇÃO $$V(aX + b) = \sigma^2_{aX+b} = a^2 \cdot \sigma^2_X \quad \text{e} \quad \sigma_{aX+b} = |a| \cdot \sigma_X$$

Esse resultado nos diz que a adição da constante b não afeta a variância, o que é intuitivo, porque a adição de b muda a localização (valor da média) mas não a dispersão dos valores. Em particular,

1. $\sigma^2_{aX} = a^2 \cdot \sigma^2_X$, $\sigma_{aX} = |a| \cdot \sigma_X$
2. $\sigma^2_{X+b} = \sigma^2_X$ (3.14)

O motivo do valor absoluto na expressão de σ_{aX} é que a pode ser negativo, ao passo que o desvio padrão não pode. a^2 resulta quando a é eliminado do termo que se eleva ao quadrado na Equação (3.13).

Exemplo 3.25

No problema da venda de computador do Exemplo 3.22, $E(X) = 2$ e

$$E(X^2) = (0)^2(0,1) + (1)^2(0,2) + (2)^2(0,3) + (3)^2(0,4) = 5$$

então $V(X) = 5 - (2)^2 = 1$. A função lucro $h(X) = 800X - 900$ tem variância $(800)^2 \cdot V(X) = (640.000)(1) = 640.000$ e desvio padrão 800. ■

Exercícios | Seção 3.3 (28–43)

28. A fmp de X = o número de defeitos graves em um eletrodoméstico selecionado aleatoriamente é

x	0	1	2	3	4
$p(x)$	0,08	0,15	0,45	0,27	0,05

Calcule os dados a seguir:

a. $E(X)$

b. $V(X)$ diretamente pela definição

c. O desvio padrão de X

d. $V(X)$ usando a fórmula alternativa.

29. Um indivíduo que possui um seguro de automóvel de uma determinada empresa é selecionado aleatoriamente. Seja Y o número de infrações ao código de trânsito para as quais o indivíduo foi reincidente nos últimos 3 anos. A fmp de Y é

y	0	1	2	3
$p(y)$	0,60	0,25	0,10	0,05

a. Calcule $E(Y)$.

b. Suponha que um indivíduo com Y infrações reincidentes incorra em uma multa de US\$ $100Y^2$. Calcule o valor esperado da multa.

30. Refira-se ao Exercício 12 e calcule $V(Y)$ e σ_Y. Determine, então, a probabilidade de Y estar dentro do intervalo de 1 desvio padrão em torno da média.

31. Uma loja de eletrodomésticos vende três modelos diferentes de *freezers* verticais com 13,5, 15,9 e 19,1 pés cúbicos de espaço, respectivamente. Seja X = volume de armazenagem comprado pelo próximo cliente a comprar um *freezer*. Suponha que a fmp de X seja

x	13,5	15,9	19,1
$p(x)$	0,2	0,5	0,3

a. Calcule $E(X)$, $E(X^2)$ e $V(X)$.

b. Se o preço de um *freezer* com X pés cúbicos de capacidade for $25X - 8,5$, qual será o preço esperado pago pelo próximo cliente a comprar um *freezer*?

c. Qual é a variância do preço $25X - 8,5$ pago pelo próximo cliente?

d. Suponha que, apesar de a capacidade nominal de um *freezer* ser X, a capacidade real seja $h(X) = X - 0,01X^2$. Qual é a capacidade real esperada do *freezer* comprado pelo próximo cliente?

32. Seja X uma va de Bernoulli com fmp como no Exemplo 3.17.

a. Calcule $E(X^2)$.

b. Demonstre que $V(X) = p(1 - p)$.

c. Calcule $E(X^{79})$.

33. Suponha que o número de certo tipo de planta encontrada em uma região retangular (denominada *quadrat* pelos ecologistas) em uma determinada área geográfica seja uma va X com fmp

$$p(x) = \begin{cases} c/x^3 & x = 1, 2, 3, \ldots \\ 0 & \text{caso contrário} \end{cases}$$

$E(X)$ é finito? Justifique sua resposta (essa é outra distribuição que os estatísticos denominam "cauda longa").

34. Uma pequena drogaria solicita cópias de uma revista para seu *display* todas as semanas. Represente por X = demanda da revista, com fmp

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{2}{15}$

Suponha que o proprietário da loja pague US\$ 1,00 por cópia e o preço de venda seja US\$ 2,00. Se as revistas não vendidas até o fim da semana não têm valor, é melhor comprar três ou quatro cópias? (*Sugestão*: Expresse a receita líquida em função da demanda X de três e quatro cópias da revista e depois calcule a receita esperada).

35. Seja X o dano (em valor monetário) incorrido por um determinado tipo de acidente em um ano. Os valores possíveis de X são 0, 1000, 5000 e 10000, com probabilidades 0,8, 0,1, 0,08 e 0,02 respectivamente. Uma empresa oferece uma apólice dedutível de US\$ 500. Se quiser que o seu lucro seja de US\$ 100, que valor de prêmio deve cobrar?

36. Os n candidatos a uma vaga foram classificados como 1, 2, 3,..., n. Seja X = a classificação de um candidato selecionado aleatoriamente, de forma que X tenha fmp

$$p(x) = \begin{cases} 1/n & x = 1, 2, 3, \ldots, n \\ 0 & \text{caso contrário} \end{cases}$$

(tal função é denominada *distribuição uniforme discreta*). Calcule $E(X)$ e $V(X)$, usando a fórmula alternativa. [*Sugestão*: a soma dos primeiros n inteiros positivos é $n(n + 1)/2$, enquanto a soma de seus quadrados é $n(n + 1)(2n + 1)/6$.]

37. Seja X = resultado de um dado lançado uma vez. Se antes de lançar o dado, oferecessem a você (1/3,5) dólares ou $h(X) = 1/X$ dólares, você aceitaria a quantia garantida ou faria a aposta? [*Nota*: Normalmente não é verdade que $1/E(X) = E(1/X)$.]

38. Uma empresa de fornecimento de materiais químicos atualmente possui em seu estoque 100 libras de um determinado produto, vendido aos clientes em lotes de 5 libras. Seja X = número de lotes pedidos por um cliente selecionado aleatoriamente e suponha que X tenha fmp

x	1	2	3	4
$p(x)$	0,2	0,4	0,3	0,1

Calcule $E(X)$ e $V(X)$ e depois o número esperado de libras restantes, após o envio do pedido do próximo cliente, e a variância do número de libras restantes. (*Sugestão*: o número de libras restantes é uma função linear de X).

39. a. Desenhe um gráfico de linhas da fmp de X do Exercício 34. Determine, então, a fmp de $-X$ e desenhe seu gráfico de linhas. Considerando as duas situações, o que você pode dizer sobre $V(X)$ e $V(-X)$?

b. Use a proposição que envolve $V(aX + b)$ para definir uma relação geral entre $V(X)$ e $V(-X)$.

40. Use a definição da Expressão (3.13) para demonstrar que $V(aX + b) = a^2 \cdot \sigma_X^2$. [*Sugestão*: com $h(X) = aX + b$, $E[h(X)] = a\mu + b$ onde $\mu = E(X)$.]

41. Suponha que $E(X) = 5$ e $E[X(X - 1)] = 27{,}5$. Qual é

a. $E(X^2)$? [*Sugestão*: $E[X(X - 1)] = E[X^2 - X] = E(X^2) - E(X)$]?

b. $V(X)$?

c. A relação geral entre as quantidades $E(X)$, $E[X(X - 1)]$ e $V(X)$?

42. Escreva a regra geral para $E(X - c)$ onde c é uma constante. O que acontece quando você assume $c = \mu$, o valor esperado de X?

43. O resultado denominado **desigualdade de Chebyshev** diz que, para qualquer distribuição de probabilidade de uma va X e qualquer número k ao menos igual a 1, $P(|X - \mu| \geq k\sigma) \leq 1/k^2$. Em palavras, a probabilidade do valor X estar a pelo menos k desvios padrão da média é de no máximo $1/k^2$.

a. Qual é o valor do limite superior para $k = 2$? $k = 3$? $k = 4$? $k = 5$? $k = 10$?

b. Calcule μ e σ da distribuição do Exercício 13. Depois calcule $P(|X - \mu| \geq k\sigma)$ para os valores de k fornecidos na parte (a). O que isto sugere sobre o limite superior em relação à probabilidade correspondente?

c. Assuma que X possa ter três valores: −1, 0, e 1, com as probabilidades $\frac{1}{18}$, $\frac{8}{9}$, e $\frac{1}{18}$, respectivamente. Qual é $P(|X - \mu| \geq 3\sigma)$, e como ela se compara ao limite correspondente?

d. Forneça uma distribuição para a qual $P(|X - \mu| \geq 5\sigma) = 0{,}04$.

3.4 Distribuição de probabilidade binomial

Há diversos experimentos que satisfazem exatamente ou aproximadamente a seguinte lista de requisitos:

1. O experimento consiste em uma seqüência de n experimentos menores denominados *tentativas*, onde n é estabelecido antes do experimento.
2. Cada tentativa pode resultar em um de dois resultados possíveis (tentativas dicotômicas), chamados de sucesso (S) ou falha (F).
3. As tentativas são independentes, de forma que o resultado de qualquer tentativa particular não influencia o resultado de qualquer outra tentativa.
4. A probabilidade de sucesso é constante de uma tentativa para a outra. Denominamos essa probabilidade p.

DEFINIÇÃO

Um experimento para o qual as Condições 1–4 são satisfeitas é denominado **experimento binomial.**

Exemplo 3.26

A mesma moeda é lançada sucessiva e independentemente n vezes. Usamos S arbitrariamente para representar o resultado H (cara) e F para representar o resultado T (coroa). Esse experimento satisfaz, então, as Condições 1–4. Lançar um percevejo n vezes, com S = ponta para cima e F = ponta para baixo, também resulta um experimento binomial. ■

Muitos experimentos envolvem uma seqüência de tentativas independentes para as quais há mais de dois resultados possíveis em qualquer tentativa. Um experimento binomial pode então ser criado, dividindo-se os resultados possíveis em dois grupos.

Exemplo 3.27

A cor das sementes de ervilhas é determinada por um único *locus* genético. Se dois alelos desse *locus* são AA ou Aa (o genótipo), então a ervilha será amarela (o fenótipo) e, se o alelo for aa, será verde. Suponha que organizemos 20 sementes Aa aos pares e cruzemos as duas ervilhas de cada par para obtenção de 10 novos genótipos. Cada novo genótipo será um sucesso S se for aa e uma falha, caso contrário. Então, com o identificador S ou F, o experimento será binomial com $n = 10$ e $p = P(\text{genótipo aa})$. Se cada membro do par for igualmente provável de contribuir com a ou A, então $p = P(\text{a}) \cdot P(\text{a}) = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$. ■

Exemplo 3.28

Suponha que certa cidade tenha 50 restaurantes licenciados, dos quais atualmente 15 possuem pelo menos uma violação grave do código de saúde e os outros 35 não possuam violações graves. Há cinco inspetores, cada um dos quais inspeciona um restaurante por semana. O nome de cada restaurante é escrito em um pedaço de papel e, após serem misturados, cada inspetor retira um dos pedaços de papel *sem reposição*. A i-ésima tentativa será um sucesso se o i-ésimo restaurante selecionado ($i = 1,..., 5$) não tiver violações graves. Então

$$P(S \text{ na primeira tentativa}) = \frac{35}{50} = 0{,}70$$

e

$$\begin{aligned} P(S \text{ na segunda tentativa}) &= P(SS) + P(FS) \\ &= P(\text{segunda } S \mid \text{primeira } S)\, P(\text{primeira } S) \\ &\quad + P(\text{segunda } S \mid \text{primeira } F)\, P(\text{primeira } F) \\ &= \frac{34}{49} \cdot \frac{35}{50} + \frac{35}{49} \cdot \frac{15}{50} = \frac{35}{50}\left(\frac{34}{49} + \frac{15}{49}\right) = \frac{35}{50} = 0{,}70 \end{aligned}$$

De forma similar, é possível mostrar que $P(S \text{ na } i\text{-ésima tentativa}) = 0{,}70$ para $i = 3, 4, 5$. Entretanto,

$$P(S \text{ na quinta tentativa} \mid SSSS) = \frac{31}{46} = 0{,}67$$

enquanto

$$P(S \text{ na quinta tentativa} \mid FFFF) = \frac{35}{46} = 0{,}76$$

O experimento não é binomial porque as tentativas não são independentes. Em geral, se a amostragem for feita sem reposição, o experimento não terá tentativas independentes. Se cada pedaço de papel fosse reposto após sua retirada, as tentativas seriam independentes, o que podia fazer com que o mesmo restaurante fosse inspecionado por mais de um inspetor. ■

Exemplo 3.29

Suponha que certo estado tenha 500.000 motoristas cadastrados, dos quais 400.000 possuem seguro. Uma amostra de 10 motoristas é selecionada sem reposição. A i-ésima tentativa será S se o i-ésimo motorista tiver

seguro. Apesar desta situação parecer idêntica à do Exemplo 3.28, a diferença importante é que o tamanho da população é muito grande em relação ao tamanho da amostra. Neste caso

$$P(S \text{ em } 2 \mid S \text{ em } 1) = \frac{399{,}999}{499{,}999} = 0{,}80000$$

e

$$P(S \text{ em } 10 \mid S \text{ no primeiro } 9) = \frac{399{,}991}{499{,}991} = 0{,}799996 \approx 0{,}80000$$

Esses cálculos sugerem que, apesar das tentativas serem exatamente independentes, as probabilidades condicionais diferem tão pouco uma da outra que, para fins práticos, as tentativas podem ser consideradas independentes com $P(S) = 0{,}8$ constante. Assim, com uma boa aproximação, o experimento é binomial com $n = 10$ e $p = 0{,}8$. ■

Usaremos a seguinte regra prática para decidir se um experimento "sem reposição" pode ser tratado como experimento binomial.

REGRA — Considere a amostragem sem reposição de uma população dicotômica de tamanho N. Se o tamanho da amostra (número de tentativas) n for no máximo 5% do tamanho da população, o experimento pode ser analisado como se fosse exatamente binomial.

Por "analisado", queremos dizer que as probabilidades baseadas nas hipóteses de experimento binomial estarão muito próximas das probabilidades reais "sem reposição", que normalmente são mais difíceis de calcular. No Exemplo 3.28, $n/N = 5/50 = 0{,}1 > 0{,}05$, de forma que o experimento binomial não é uma boa aproximação, mas, no Exemplo 3.29, $n/N = 10/500.000 < 0{,}05$.

Variável aleatória binomial e sua distribuição

Na maioria dos experimentos binomiais, interessa o número total de S e não o conhecimento de exatamente quais tentativas resultaram em S.

DEFINIÇÃO — Dado um experimento binomial consistindo de n tentativas, a **variável aleatória binomial** X a ele associada é definida como

$$X = \text{quantidade de } S \text{ nas } n \text{ tentativas}$$

Suponha, por exemplo, que $n = 3$. Haverá, então, oito resultados possíveis para o experimento:

SSS SSF SFS SFF FSS FSF FFS FFF

A partir da definição de X, $X(SST) = 2$, $X(SFF) = 1$ e assim por diante. Os valores possíveis de X em um experimento de n tentativas são $x = 0, 1, 2, ..., n$. Freqüentemente usaremos $X \sim \text{Bin}(n, p)$ para indicar que X é uma va binomial baseada em n tentativas com probabilidade de sucesso p.

NOTAÇÃO — Como a fmp de uma va binomial depende de dois parâmetros n e p, representamos a fmp por $b(x; n, p)$.

Considere primeiro o caso de $n = 4$ para o qual cada resultado, sua probabilidade e o valor de x correspondente estão relacionados na Tabela 3.1. Por exemplo:

$$\begin{aligned} P(SSFS) &= P(S) \cdot P(S) \cdot P(F) \cdot P(S) && \text{(tentativas independentes)} \\ &= p \cdot p \cdot (1 - p) \cdot p && (P(S) \text{ constante}) \\ &= p^3 \cdot (1 - p) \end{aligned}$$

Tabela 3.1 Resultados e probabilidades de um experimento binominal com quatro tentativas

Resultado	x	Probabilidade	Resultado	x	Probabilidade
$SSSS$	4	p^4	$FSSS$	3	$p^3(1-p)$
$SSSF$	3	$p^3(1-p)$	$FSSF$	2	$p^2(1-p)^2$
$SSFS$	3	$p^3(1-p)$	$FSFS$	2	$p^2(1-p)^2$
$SSFF$	2	$p^2(1-p)^2$	$FSFF$	1	$p(1-p)^3$
$SFSS$	3	$p^3(1-p)$	$FFSS$	2	$p^2(1-p)^2$
$SFSF$	2	$p^2(1-p)^2$	$FFSF$	1	$p(1-p)^3$
$SFFS$	2	$p^2(1-p)^2$	$FFFS$	1	$p(1-p)^3$
$SFFF$	1	$p(1-p)^3$	$FFFF$	0	$(1-p)^4$

Neste caso especial, queremos $b(x; 4, p)$ para $x = 0, 1, 2, 3$ e 4. Para $b(3; 4, p)$, identificamos quais dos 16 resultados fornecem um valor de x igual a 3 e somamos as probabilidades associadas a cada resultado:

$$b(3; 4, p) = P(FSSS) + P(SFSS) + P(SSFS) + P(SSSF) = 4p^3(1 - p)$$

Há quatro resultados com $x = 3$, cada um com probabilidade $p^3(1 - p)$ (a ordem de S e F não é importante, apenas a quantidade S), então

$$b(3; 4, p) = \begin{Bmatrix} \text{número de resultados} \\ \text{com } X = 3 \end{Bmatrix} \cdot \begin{Bmatrix} \text{probabilidade de qualquer resultado} \\ \text{particular com } X = 3 \end{Bmatrix}$$

De forma similar, $b(2; 4, p) = 6p^2(1 - p)^2$, que também é o produto do número de resultados com $X = 2$ e a probabilidade qualquer de tais resultados. Em geral,

$$b(x; n, p) = \begin{Bmatrix} \text{número de seqüências de} \\ \text{comprimento } n \text{ consistindo de } x\ S's \end{Bmatrix} \cdot \begin{Bmatrix} \text{probabilidade de qualquer resultado} \\ \text{particular de tal seqüência} \end{Bmatrix}$$

Como a ordenação de S e F não é importante, o segundo fator da equação anterior é $p^x(1 - p)^{n-x}$ (por exemplo: as primeiras x tentativas resultarem em S e as últimas $n - x$ resultarem em F). O primeiro fator é o número de maneiras de escolher x das n tentativas serem S, isto é, o número de combinações de tamanho x que podem ser construídas com n objetos distintos (tentativas, no caso).

TEOREMA

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & x = 0, 1, 2, \ldots n \\ 0 & \text{caso contrário} \end{cases}$$

Exemplo 3.30

Cada um de seis consumidores de refrigerante selecionados aleatoriamente recebe um copo com o refrigerante S e um com o refrigerante F. Os copos são idênticos, exceto por um código no fundo que identifica o refrigerante. Suponha que não haja uma tendência de preferência entre os consumidores. Então $p = P$(um indivíduo selecionado prefere S) $= 0{,}5$, de forma que X = número de consumidores entre os seis que preferem S, $X \sim \text{Bin}(6, 0{,}5)$.

Dessa forma,

$$P(X = 3) = b(3; 6, 0,5) = \binom{6}{3}(0,5)^3(0,5)^3 = 20(0,5)^6 = 0,313$$

A probabilidade de pelo menos três preferirem S é

$$P(3 \leq X) = \sum_{x=3}^{6} b(x; 6, 0,5) = \sum_{x=3}^{6} \binom{6}{x}(0,5)^x(0,5)^{6-x} = 0,656$$

e a probabilidade de no máximo um preferir S é

$$P(X \leq 1) = \sum_{x=0}^{1} b(x; 6, 0,5) = 0,109$$ ■

Uso das tabelas binomiais

Mesmo para um valor relativamente pequeno de n, o cálculo da probabilidade binomial é tedioso. A Tabela do Apêndice A.1 exibe a FDA $F(x) = P(X \leq x)$ para $n = 5, 10, 15, 20, 25$ em combinação com determinados valores de p. Diversas outras probabilidades podem ser calculadas, usando-se a proposição de FDA da Seção 3.2.

NOTAÇÃO Para X (distribuída como (o sinal é um Til, que não significa aproximadamente)) $\sim \text{Bin}(n, p)$, a FDA será representada por

$$P(X \leq x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p) \qquad x = 0, 1, \ldots, n$$

Exemplo 3.31

Suponha que 20% de todas as cópias de um livro-texto apresentem falha em um determinado teste de resistência de encadernação. Seja X o número de cópias que apresentam falhas entre 15 cópias selecionadas aleatoriamente. Então X tem distribuição binomial com $n = 15$ e $p = 0,2$.

1. A probabilidade de no máximo 8 apresentarem falha é

$$P(X \leq 8) = \sum_{y=0}^{8} b(y; 15, 0,2) = B(8; 15, 0,2)$$

que é a entrada na linha $x = 8$ e coluna $p = 0,2$ da tabela binomial $n = 15$. Na Tabela do Apêndice A.1, a probabilidade é $B(8; 15, 0,2) = 0,999$.

2. A probabilidade de exatamente 8 apresentarem falha é

$$P(X = 8) = P(X \leq 8) - P(X \leq 7) = B(8; 15, 0,2) - B(7; 15, 0,2)$$

que é a diferença entre duas entradas consecutivas na coluna $p = 0,2$. O resultado é $0,999 - 0,996 = 0,003$.

3. A probabilidade de no mínimo 8 apresentarem falha é

$$P(X \geq 8) = 1 - P(X \leq 7) = 1 - B(7; 15, 0,2)$$

$$= 1 - \begin{pmatrix} \text{entrada em } x = 7 \\ \text{linha da coluna } p = 0,2 \end{pmatrix}$$

$$= 1 - 0,996 = 0,004$$

4. Finalmente, a probabilidade de 4 a 7, inclusive, apresentarem falha é

$$P(4 \leq X \leq 7) = P(X = 4, 5, 6 \text{ ou } 7) = P(X \leq 7) - P(X \leq 3)$$
$$= B(7; 15, 0{,}2) - B(3; 15, 0{,}2) = 0{,}996 - 0{,}648 = 0{,}348$$

Observe que a última probabilidade é a diferença entre as entradas nas linhas $x = 7$ e $x = 3$ e *não* das linhas $x = 7$ e $x = 4$. ■

Exemplo 3.32

Um fabricante de eletrônicos alega que no máximo 10% de seus geradores precisam de reparo no período de garantia. Para investigar a declaração, técnicos de um laboratório de teste compraram 20 unidades e as submeteram a um teste acelerado para simular o uso durante o período de garantia. Seja p a probabilidade de um gerador precisar de reparo durante o período (proporção de unidades que necessitam de reparo). Os técnicos do laboratório devem decidir se os dados resultantes do experimento suportam a alegação de $p \leq 0{,}10$. Seja X o número entre os 20 da amostra que precisam de reparo, de forma que $X \sim \text{Bin}(20, p)$. Considere a regra de decisão.

Rejeitar a alegação de $p \leq 0{,}10$ em favor da conclusão de que $p > 0{,}10$ se $x \geq 5$ (onde x é o valor observado de X) e considerar a alegação aceitável se $x \leq 4$.

A probabilidade de a alegação ser rejeitada quando $p = 0{,}10$ (conclusão incorreta) é

$$P(X \geq 5 \text{ quando } p = 0{,}10) = 1 - B(4; 20, 0{,}1) = 1 - 0{,}957 = 0{,}043$$

A probabilidade de a alegação não ser rejeitada quando $p = 0{,}20$ (tipo diferente de conclusão incorreta) é

$$P(X \leq 4 \text{ quando } p = 0{,}2) = B(4; 20, 0{,}2) = 0{,}630$$

A primeira probabilidade é relativamente pequena, mas a segunda é intoleravelmente grande. Quando $p = 0{,}20$, o que significa que o fabricante errou grosseiramente para menos no percentual de unidades que necessitam de reparo, e a regra de decisão estabelecida for usada, 63% de todas as amostras serão incluídas na declaração do fabricante ser julgada aceitável!

Pode-se pensar que a probabilidade deste segundo tipo de conclusão errônea poderia ser menor se o valor de corte 5 fosse alterado para outro valor na regra de decisão. Entretanto, apesar de a substituição de 5 por um número menor resultar em uma probabilidade menor que 0,630, a outra probabilidade aumentaria. A única forma de diminuir as duas "probabilidades de erro" é basear a regra de decisão em um experimento que envolva muito mais unidades. ■

Observe que um valor tabelado igual a 0 significa apenas que a probabilidade é 0 com três dígitos de significativos, pois todas as entradas da tabela são positivas. Programas estatísticos como o MINITAB geram $b(x; n, p)$ ou $B(x; n, p)$ para quaisquer valores especificados de n e p. No Capítulo 4, apresentaremos um método rápido e preciso de aproximações de probabilidades binomiais quando n for grande.

Média e variância de X

Para $n = 1$, a distribuição binomial se torna a distribuição de Bernoulli. Pelo Exemplo 3.17, o valor médio de uma variável Bernoulli é $\mu = p$, de forma que o número esperado de S em uma tentativa única é p. Como um experimento binomial consiste de n tentativas, a intuição sugere que, para $X \sim \text{Bin}(n, p)$, $E(X) = np$, o produto do número de tentativas pela probabilidade de sucesso em uma única tentativa. A expressão de $V(X)$ não é tão intuitiva.

PROPOSIÇÃO Se $X \sim \text{Bin}(n, p)$, então $E(X) = np$, $V(X) = np(1 - p) = npq$, e $\sigma_X = \sqrt{npq}$ (onde $q = 1 - p$).

Dessa forma, o cálculo da média e da variância de uma va binomial não requer o cálculo de somatórias. A demonstração do resultado de $E(X)$ está esquematizada no Exercício 60.

Exemplo 3.33

Se 75% de todas as compras em uma determinada loja forem feitas com cartão de crédito e X for a quantidade de compras feitas com cartão de crédito entre 10 compras selecionadas aleatoriamente, então $X \sim \text{Bin}(10; 0{,}75)$. Portanto, $E(X) = np = (10)(0{,}75) = 7{,}5$, $V(X) = npq = 10(0{,}75)(0{,}25) = 1{,}875$, e $\sigma = \sqrt{1{,}875}$. Novamente, apesar de X só poder assumir valores inteiros, $E(X)$ não precisa ser um inteiro. Se executarmos um grande número de experimentos binomiais independentes, cada um com $n = 10$ tentativas e $p = 0{,}75$, o número médio de S por experimento será próximo de 7,5. ■

Exercícios | Seção 3.4 (44–63)

44. Calcule as seguintes probabilidades binomiais diretamente pela fórmula de $b(x; n, p)$:

a. $b(3; 8, 0{,}6)$

b. $b(5; 8, 0{,}6)$

c. $P(3 \leq X \leq 5)$ quando $n = 8$ e $p = 0{,}6$

d. $P(1 \leq X)$ quando $n = 12$ e $p = 0{,}1$

45. Use a Tabela do Apêndice A.1 para obter as probabilidades a seguir:

a. $B(4; 10, 0{,}3)$

b. $b(4; 10, 0{,}3)$

c. $b(6; 10, 0{,}7)$

d. $P(2 \leq X \leq 4)$ quando $X \sim \text{Bin}(10, 0{,}3)$

e. $P(2 \leq X)$ quando $X \sim \text{Bin}(10, 0{,}3)$

f. $P(X \leq 1)$ quando $X \sim \text{Bin}(10, 0{,}7)$

g. $P(2 < X < 6)$ quando $X \sim \text{Bin}(10, 0{,}3)$

46. Quando as placas de circuito integrado usadas na fabricação de CD-players são testadas, a porcentagem de placas com defeitos no longo prazo é igual a 5%. Seja X = número de placas com defeito em uma amostra aleatória de tamanho $n = 25$, de forma que $X \sim \text{Bin}(25, 0{,}05)$.

a. Determine $P(X \leq 2)$.

b. Determine $P(X \geq 5)$.

c. Determine $P(1 \leq X \leq 4)$.

d. Qual é a probabilidade de que nenhuma das 25 placas apresente defeito?

e. Calcule o valor esperado e o desvio padrão de X.

47. Uma empresa de cristais finos sabe por experiência que 10% de suas taças possuem defeitos cosméticos e devem ser classificadas como "de segunda linha".

a. Entre seis taças selecionadas aleatoriamente, qual é a probabilidade de uma ser de segunda linha?

b. Entre seis taças selecionadas aleatoriamente, qual é a probabilidade de no mínimo duas serem de segunda linha?

c. Se as taças forem examinadas uma a uma, qual será a probabilidade de no máximo cinco terem de ser selecionadas para encontrar quatro que não sejam de segunda linha?

48. Suponha que apenas 25% de todos os motoristas parem completamente um cruzamento com semáforos vermelhos para todas as direções quando não há outros carros à vista. Qual é a probabilidade de que, entre 20 motoristas selecionados aleatoriamente chegando em um cruzamento nessas condições,

a. no máximo seis parem totalmente?

b. exatamente seis parem completamente?

c. ao menos seis parem completamente?

d. quantos dos 20 motoristas você espera que parem completamente?

49. O Exercício 29 (Seção 3.3) forneceu a fmp de Y, o número de violações do código de trânsito de um indivíduo selecionado aleatoriamente com seguro de uma certa empresa. Qual é a probabilidade de que entre 15 de tais indivíduos selecionados aleatoriamente,

a. pelo menos 10 não tenham violações?

b. menos da metade tenha ao menos uma violação?

c. o número das pessoas com ao menos uma violação esteja entre 5 e 10, inclusive?[2]

50. Um determinado tipo de raquete de tênis possui duas versões: média e grande. Sessenta por cento de todos os clientes de certa loja querem a versão grande.

a. Entre 10 clientes selecionados aleatoriamente que querem esse tipo de raquete, qual é a probabilidade de ao menos seis quererem a versão grande?

b. Entre 10 clientes selecionados aleatoriamente, qual é a probabilidade de o número dos que desejam a versão grande estar dentro do intervalo de um desvio padrão da média?

c. A loja possui atualmente sete raquetes de cada versão. Qual é a probabilidade de os 10 próximos clientes que querem essa raquete conseguirem obter a versão desejada do estoque atual?

51. Vinte por cento de todos os telefones de um determinado tipo são enviados para reparo em garantia. Desses, 60% podem ser reparados, enquanto os outros 40%

[2] "Entre a e b, inclusive" é equivalente a $(a \leq X \leq b)$.

devem ser substituídos. Se uma empresa comprar 10 desses telefones, qual é a probabilidade de exatamente dois serem substituídos em garantia?

52. O College Board relata que 2% dos 2 milhões de alunos que fazem o SAT cada ano recebem acomodações especiais por causa de deficiência física documentada (*Los Angeles Times*, July 16, 2002). Considere uma amostra aleatória de 25 estudantes que fizeram o teste recentemente.

a. Qual é a probabilidade de exatamente 1 ter recebido acomodação especial?

b. Qual é a probabilidade de ao menos 1 ter recebido acomodação especial?

c. Qual é a probabilidade de ao menos 2 terem recebido acomodação especial?

d. Qual é a probabilidade de o número entre os 25 que receberam acomodação especial estar dentro de 2 desvios padrão do número que você espera serem acomodados?

e. Suponha que um estudante que não receba acomodação especial tenha 3 horas para fazer o exame e um que receba tenha 4,5. Qual você espera que seja o tempo médio dos 25 estudantes selecionados?

53. Suponha que 90% de todas as pilhas de certo fabricante tenham voltagens aceitáveis. Um determinado tipo de lanterna necessita de duas pilhas tipo D, e ela só funciona se as duas pilhas tiverem voltagem aceitável. Entre 10 lanternas selecionadas aleatoriamente, qual é a probabilidade de pelo menos nove funcionarem? Que hipótese você fez no decorrer da resposta à pergunta proposta?

54. Um grande lote de componentes chegou em um distribuidor e só pode ser classificado como aceitável se a proporção de componentes com defeito for no máximo 0,10. O distribuidor decide selecionar 10 componentes aleatoriamente e aceitar o lote apenas se o número de componentes defeituosos da amostra for no máximo 2.

a. Qual é a probabilidade de o lote ser aceito quando a proporção real de itens com defeito for 0,01? 0,05? 0,10? 0,20? 0,25?

b. Seja p a proporção real de itens com defeito no lote. Um gráfico de *P*(lote é aceito) em função de p, com p no eixo horizontal e *P*(lote é aceito) no eixo vertical é denominado *curva característica de operação* do plano de aceitação da amostragem. Use os resultados da parte (a) para fazer o gráfico da curva para $0 \leq p \leq 1$.

c. Repita as partes (a) e (b) com "1" substituindo "2" no plano de aceitação da amostragem.

d. Repita as partes (a) e (b) com "15" substituindo "10" no plano de aceitação da amostragem.

e. Qual dos três planos de amostragem é mais satisfatório e por quê?

55. Uma norma que exige a instalação de um detector de fumaça em todas as casas construídas anteriormente está em vigor há um ano em certa cidade. O corpo de bombeiros está preocupado porque muitas casas continuam sem detectores. Seja p = proporção real de casas que possuem detectores e suponha que uma amostra aleatória de 25 lares seja inspecionada. Se a amostra indica fortemente que pouco menos de 80% de todas as casas possuem detector, o corpo de bombeiros fará uma campanha por um programa de inspeção obrigatória. Por causa do custo do programa, o corpo de bombeiros prefere não pedir essas inspeções a menos que a amostra renda evidências fortes que comprovem sua necessidade. Seja X o número de casas com detectores entre as 25 da amostra. Considere rejeitar a hipótese de $p \geq 0,8$ se $x \leq 15$.

a. Qual é a probabilidade de a hipótese ser rejeitada se o valor real de p for 0,8?

b. Qual é a probabilidade de não rejeitar a hipótese quando $p = 0,7$? Quando $p = 0,6$?

c. Como as "probabilidades de erro" das partes (a) e (b) mudam se o valor 15 na regra de decisão for alterado para 14?

56. Uma ponte cobra um pedágio de US$ 1,00 para carros de passeio e de US$ 2,50 para outros veículos. Suponha que, durante o dia, 60% de todos os veículos sejam carros de passeio. Se 25 veículos cruzarem a ponte durante um determinado período do dia, qual será a receita esperada resultante? [*Sugestão:* Seja X = número de carros de passeio. A receita total $h(X)$ é uma função linear de X.]

57. Um estudante que está tentando escrever um trabalho para um curso tem a escolha de dois tópicos: A e B. Se o aluno escolher o tópico A, solicitará dois livros por empréstimo da biblioteca, e, se escolher B, serão solicitados quatro livros. O estudante acredita que, para escrever um bom trabalho, precisa receber e usar ao menos metade dos livros selecionados para cada tópico escolhido. Se a probabilidade de um livro solicitado chegar em tempo for de 0,9 e os livros chegam independentemente um do outro, que tópico o aluno deve escolher para maximizar a probabilidade de escrever um bom artigo? E se a probabilidade for de apenas 0,5 em vez de 0,9?

58. a. Para n determinado, há valores de p $(0 \leq p \leq 1)$ para os quais $V(X) = 0$? Explique.

b. Para que valor de p $V(X)$ é maximizada? [*Sugestão:* faça o gráfico de $V(X)$ em função de p ou calcule a derivada.]

59. a. Demonstre que $b(x; n, 1-p) = b(n-x; n, p)$.

b. Demonstre que $B(x; n, 1-p) = 1 - B(n-x-1; n, p)$. [*Sugestão:* no máximo x *S's* serão equivalentes a pelo menos $(n-x)$ *F's*.]

c. O que as partes (a) e (b) implicam sobre a necessidade de se incluir valores de p maiores do que 0,5 na Tabela A.1 do Apêndice?

60. Demonstre que $E(X) = np$ quando X é uma variável aleatória binomial. [*Sugestão*: primeiro expresse $E(X)$ como uma soma com limite inferior $x = 1$. Depois fatore

np, assumindo $y = x - 1$ de forma que a soma varie de $y = 0$ a $y = n - 1$ e mostre que a soma é igual a 1.]

61. Os clientes de um posto de gasolina pagam com cartão de crédito (A), cartão de débito (B) ou dinheiro (C). Assuma que clientes sucessivos façam escolhas independentes, com $P(A) = 0{,}5$, $P(B) = 0{,}2$ e $P(C) = 0{,}3$.

a. Entre os próximos 100 clientes, qual será a média e a variância do número dos que pagam com cartão de débito? Explique seu raciocínio.

b. Responda à parte (a) para o número entre os 100 que não pagam em dinheiro.

62. Uma limusine de aeroporto pode acomodar até quatro passageiros em qualquer corrida. A empresa aceitará um máximo de seis reservas e os passageiros devem ter reservas. Pelos registros anteriores, 20% de todos os que fazem reservas não aparecem para a corrida. Responda as seguintes perguntas, assumindo independência quando apropriado.

a. Se forem feitas seis reservas, qual é a probabilidade de ao menos um indivíduo com reserva não poder ser acomodado na corrida?

b. Se forem feitas seis reservas, qual é o número esperado de lugares disponíveis quando a limusine parte?

c. Suponha que a distribuição de probabilidade do número de reservas feitas seja dada na tabela a seguir.

Número de reservas	3	4	5	6
Probabilidade	0,1	0,2	0,3	0,4

Seja X = número de passageiros de uma corrida selecionada aleatoriamente. Calcule a função de distribuição de probabilidade de X.

63. Refira-se à desigualdade de Chebyshev fornecida no Exercício 43. Calcule $P(|X - \mu| \geq k\sigma)$ para $k = 2$ e $k = 3$ quando $X \sim \text{Bin}(20, 0{,}5)$ e compare com o limite superior correspondente. Repita para $X \sim \text{Bin}(20, 0{,}75)$.

3.5 Distribuições binomiais hipergeométrica e negativa

As distribuições binomiais hipergeométrica e negativa são ambas intimamente relacionadas à distribuição binomial. Enquanto a distribuição binomial é o modelo aproximado de amostragem sem reposição de uma população (*S-F*) dicotômica finita, a distribuição hipergeométrica é o modelo de probabilidade para o número de *S's* em uma amostra. A va binomial X é o número de S quando o número de tentativas n está estabelecido, enquanto a distribuição binomial negativa resulta da fixação do número de S desejado, deixando que o número de tentativas seja aleatório.

Distribuição hipergeométrica

As hipóteses que levam à distribuição hipergeométrica são as seguintes:

1. A população ou o conjunto de onde é retirada a amostra consiste de N indivíduos, objetos ou elementos (população *finita*).
2. Cada indivíduo é classificado como sucesso (S) ou falha (F) e há M sucessos na população.
3. É selecionada uma amostra sem reposição de n indivíduos de forma que cada subconjunto de tamanho n seja igualmente provável de ser escolhido.

A variável aleatória de interesse é X = número de *S's* na amostra. A distribuição de probabilidade de X depende dos parâmetros n, M e N, de forma que queremos calcular $P(X = x) = h(x; n, M, N)$.

Exemplo 3.34

Durante determinado período, um escritório de tecnologia da informação de uma universidade recebeu 20 ordens de serviço de problemas com impressoras, das quais 8 de impressoras a laser e 12 a jato de tinta. Uma amostra de 5 dessas ordens de serviço será selecionada para inclusão em uma pesquisa de satisfação do cliente. Suponha que as 5 sejam selecionadas de forma completamente aleatória para que qualquer subconjunto de tamanho 5 tenha a mesma possibilidade de ser selecionado (imagine colocar os números 1, 2,..., 20 em 20 tarjas de papel, misturá-las e escolher cinco delas). Qual será a probabilidade de exatamente x ($x = 0, 1, 2, 3, 4$ ou 5) das ordens de serviço selecionadas serem de impressoras a jato de tinta?

Neste exemplo, o tamanho da população é $N = 20$, o tamanho da amostra é $n = 5$ e o número de S (jato de tinta = S) e F da população são $M = 12$ e $N - M = 8$, respectivamente. Considere o valor $x = 2$. Como todos os resultados (cada um consistindo em 5 ordens) são igualmente prováveis,

$$P(X = 2) = h(2; 5, 12, 20) = \frac{\text{número de resultados com } X = 2}{\text{número de resultados possíveis}}$$

O número de resultados possíveis no experimento é o número de maneiras de selecionar 5 dos 20 objetos sem se importar com a ordem, isto é, $\binom{20}{5}$. Para contar o número de resultados tendo $X = 2$, observe que há $\binom{12}{2}$ formas de selecionar 2 das ordens de serviço de impressoras a jato de tinta e para cada forma há $\binom{8}{3}$ maneiras de selecionar as 3 ordens de serviço de impressoras a laser para completar a amostra. A regra do produto do Capítulo 2 fornece então $\binom{12}{2}\binom{8}{3}$ como o número de resultados com $X = 2$, assim

$$h(2; 5, 12, 20) = \frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}} = \frac{77}{323} = 0{,}238 \qquad \blacksquare$$

Em geral, se o tamanho da amostra n for menor que o número de sucessos da população (M), o maior valor possível de X será n. Entretanto, se $M < n$ (por exemplo, uma amostra de tamanho 25 e apenas 15 sucessos na população), então X pode ser no máximo M. De forma similar, sempre que o número de falhas da população ($N - M$) exceder o tamanho da amostra, o menor valor possível de X será 0 (porque todos os indivíduos da amostra podem ser falhas). Entretanto, se $N - M < n$, o menor valor possível de X será $n - (N - M)$. Resumindo, os valores possíveis de X satisfazem à restrição máx.$(0, n - (N - M)) \leq x \leq$ mín.(n, M). Um argumento paralelo ao do exemplo anterior fornece a fmp de X.

PROPOSIÇÃO Se X for o número de S de uma amostra completamente aleatória de tamanho n tirada de uma população constituída de M S's e $(N - M)$ F's, então a distribuição de probabilidade de X, denominada **distribuição hipergeométrica,** será dada por

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad (3.15)$$

para um inteiro x que satisfaça máx.$(0, n - N + M) \leq x \leq$ mín.(n, M).

No Exemplo 3.34, $n = 5$, $M = 12$ e $N = 20$, de forma que $h(x; 5, 12, 20)$ para $x = 0, 1, 2, 3, 4, 5$ pode ser obtida, substituindo-se esses números na Equação (3.15).

Exemplo 3.35

Cinco indivíduos de uma população animal supostamente ameaçada de extinção em certa região foram capturados, marcados e liberados para se misturarem à população. Após terem uma oportunidade de cruzarem, foi selecionada uma amostra aleatória de 10 desses animais. Seja X = número de animais marcados na segunda amostra. Se, na verdade, houver 25 animais desse tipo na região, qual será a probabilidade de (a) $X = 2$? (b) $X \leq 2$?

Os valores dos parâmetros são $n = 10$, $M = 5$ (5 animais marcados na população) e $N = 25$, assim

$$h(x; 10, 5, 25) = \frac{\binom{5}{x}\binom{20}{10-x}}{\binom{25}{10}} \qquad x = 0, 1, 2, 3, 4, 5$$

Para a parte (a),

$$P(X = 2) = h(2; 10, 5, 25) = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} = 0{,}385$$

Para a parte (b),

$$P(X \leq 2) = P(X = 0, 1, \text{ ou } 2) = \sum_{x=0}^{2} h(x; 10, 5, 25)$$
$$= 0{,}057 + 0{,}257 + 0{,}385 = 0{,}699$$

■

Estão disponíveis tabelas abrangentes da distribuição hipergeométrica, mas, como a distribuição possui três parâmetros, as tabelas requerem muito mais espaço do que as da distribuição binomial. O MINITAB e outros programas estatísticos geram probabilidades hipergeométricas facilmente.

Como no caso binomial, há expressões simples para $E(X)$ e $V(X)$ para vas hipergeométricas.

PROPOSIÇÃO A média e a variância da va hipergeométrica X com fmp $h(x; n, M, N)$ são

$$E(X) = n \cdot \frac{M}{N} \qquad V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

A razão M/N é a proporção de S na população. Se substituirmos M/N por p em $E(X)$ e $V(X)$, obtemos

$$E(X) = np$$
$$V(X) = \left(\frac{N-n}{N-1}\right) \cdot np(1-p) \tag{3.16}$$

A Expressão (3.16) mostra que as médias das vas binomial e hipergeométrica são iguais, enquanto as variâncias das duas vas diferem pelo fator $(N - n)/(N - 1)$, freqüentemente denominado **fator de correção de população finita.** Esse fator é menor do que 1, de forma que a variável hipergeométrica tem variância menor do que a va binomial. O fator de correção pode ser escrito como $(1 - n/N)/(1 - 1/N)$, que é aproximadamente 1 quando n é pequeno em relação a N.

Exemplo 3.36 (Continuação do Exemplo 3.35)

No exemplo de marcação dos animais, $n = 10$, $M = 5$, e $N = 25$, de forma que $p = \frac{5}{25} = 0{,}2$ e

$$E(X) = 10(0{,}2) = 2$$

$$V(X) = \frac{15}{24}(10)(0{,}2)(0{,}8) = (0{,}625)(1{,}6) = 1$$

Se a amostragem fosse feita com reposição, $V(X) = 1{,}6$.

Suponha que o tamanho da população N não seja conhecido, de forma que o valor x é observado e queremos estimar N. É razoável igualar a proporção da amostra observada de S, x/n, à proporção da população, M/N, fornecendo a estimativa

$$\hat{N} = \frac{M \cdot n}{x}$$

Se $M = 100$, $n = 40$ e $x = 16$, então $\hat{N} = 250$. ■

Nossa regra prática geral da Seção 3.4 afirmava que, se a amostragem foi feita sem reposição mas n/N era no máximo 0,05, a distribuição binomial podia ser usada para calcular probabilidades aproximadas, envolvendo o número de S na amostra. Segue uma definição mais precisa: considere que o tamanho da população N e o número S's da população M, cresçam à medida que a relação M/N aproxima-se de p. Então $h(x; n, M, N)$ se aproxima de $b(x; n, p)$. Para n/N pequeno, os dois são aproximadamente iguais, desde que p não esteja muito perto tanto de 0 como de 1. Essa é a base de nossa regra prática.

Distribuição Binomial Negativa

A va e a distribuição binomial negativa são baseadas em experimentos que satisfaçam às condições a seguir:

1. O experimento consiste de uma seqüência de tentativas independentes.

2. Cada tentativa resulta em sucesso (S) ou em falha (F).

3. A probabilidade de sucesso é constante de uma tentativa para outra, então $P(S$ na tentativa $i) = p$ para $i = 1,2,3....$

4. O experimento continua (as tentativas são executadas) até ser observado um total de r sucessos, sendo r um inteiro positivo.

A variável aleatória de interesse é X = número de falhas que precedem o r-ésimo sucesso. X é denominada **variável aleatória binomial negativa** porque, em contraste com a va binomial, o número de sucessos é fixo e o de tentativas é aleatório.

Valores possíveis de X são 0, 1, 2, ... Represente por $nb(x; r, p)$ a fmp de X. O evento $\{X = x\}$ é equivalente a $\{r - 1\ S$'s nas primeiras $(x + r - 1)$ tentativas e um S na $(x + r)$*ésima* tentativa$\}$ (por exemplo: se $r = 5$ e $x = 10$, deve haver quatro S nas primeiras 14 tentativas e a tentativa 15 deve ser um S). Como as tentativas são independentes,

$$\begin{aligned} nb(x; r, p) &= P(X = x) \\ &= P(r - 1\ S\text{'s nas primeiras } x + r - 1 \text{ tentativas}) \cdot P(S) \end{aligned} \qquad (3.17)$$

A primeira probabilidade do membro direito da Expressão (3.17) é a probabilidade binomial

$$\binom{x + r - 1}{r - 1} p^{r-1}(1 - p)^x \qquad \text{onde } P(S) = p$$

PROPOSIÇÃO

> A fmp da va binomial negativa X com parâmetros r = número de S's e $p = P(S)$ é
>
> $$nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r(1 - p)^x \qquad x = 0, 1, 2, \ldots$$

Exemplo 3.37

Um pediatra deseja convocar cinco casais, cada um esperando seu primeiro filho, para participarem de um novo regime de parto natural. Seja $p = P$ (um casal selecionado aleatoriamente concorda em participar). Se $p = 0{,}2$, qual é a probabilidade de 15 casais serem solicitados antes de serem encontrados cinco que concordem em participar? Isto é, com $S = \{$concorda em participar$\}$, qual é a probabilidade de ocorrerem 10 F antes do quinto S? Substituindo $r = 5$, $p = 0{,}2$ e $x = 10$ em $nb(x; r, p)$, temos

$$nb(10; 5, 0{,}2) = \binom{14}{4}(0{,}2)^5(0{,}8)^{10} = 0{,}034$$

A probabilidade de no máximo 10 F serem observados (no máximo 15 casais serem entrevistados) é

$$P(X \leq 10) = \sum_{x=0}^{10} nb(x; 5, 0{,}2) = (0{,}2)^5 \sum_{x=0}^{10} \binom{x+4}{4} (0{,}8)^x = 0{,}164 \quad ■$$

Em algumas fontes de informações, a va binomial negativa é considerada como o número de tentativas $X + r$ em vez de número de falhas.

No caso particular de $r = 1$, a fmp é

$$nb(x; 1, p) = (1 - p)^x p \qquad x = 0, 1, 2, \ldots \qquad (3.18)$$

No Exemplo 3.10, deduzimos a fmp em relação ao número de tentativas necessárias para se obter o primeiro S e aquela fmp é semelhante à Expressão (3.18). Tanto X = número de F como Y = número de tentativas ($= 1 + X$) são denominadas **variáveis aleatórias geométricas** na literatura especializada e a fmp da Expressão (3.18) é denominada **distribuição geométrica.**

No Exemplo 3.18, foi demonstrado que o valor esperado de tentativas até a obtenção do primeiro S é $1/p$, de forma que o valor esperado de F até a obtenção do primeiro S é $(1/p) - 1 = (1 - p)/p$. Intuitivamente, esperaríamos ver $r \cdot (1 - p)/p$ F antes do r-ésimo S e esta é, na verdade, $E(X)$. Também há uma fórmula simples para $V(X)$.

PROPOSIÇÃO

Se X é uma va binomial negativa com fmp $nb(x; r, p)$, então

$$E(X) = \frac{r(1-p)}{p} \qquad V(X) = \frac{r(1-p)}{p^2}$$

Finalmente, desenvolvendo o coeficiente binomial de $p^r(1 - p)^x$ e fazendo algumas simplificações, vê-se que $nb(x; r, p)$ é bem definido, mesmo quando r não é inteiro. Verifica-se que a *distribuição binomial negativa geral* ajusta muito bem os dados observados em uma grande variedade de aplicações.

Exercícios Seção 3.5 (64–74)

64. Certo tipo de câmera digital é oferecida em duas versões de três megapixel e quatro megapixel. Uma loja de câmeras recebeu uma encomenda de 15 dessas câmeras, das quais seis com resolução de três megapixel. Suponha que cinco delas sejam selecionadas aleatoriamente para serem estocadas atrás do balcão. As outras 10 são colocadas na área de armazenagem. Seja X = número de câmeras de três megapixel entre as cinco selecionadas para armazenagem atrás do balcão.

a. Que tipo de distribuição tem X (nome e valores de todos os parâmetros)?

b. Calcule $P(X = 2)$, $P(X \leq 2)$ e $P(X \geq 2)$.

c. Calcule o valor médio e o desvio padrão de X.

65. Doze refrigeradores de um determinado tipo foram devolvidos ao distribuidor por causa de um ruído audível, oscilante e agudo que faziam quando em funcionamento. Suponha que sete desses refrigeradores possuam um compressor defeituoso e os outros cinco tenham problemas graves. Se os refrigeradores forem examinados em ordem aleatória, seja X o número de refrigeradores examinados entre os seis primeiros com compressores defeituosos. Calcule os seguintes valores:

a. $P(X = 5)$

b. $P(X \leq 4)$

c. A probabilidade de X exceder o valor da média em mais de 1 desvio padrão.

d. Considere uma entrega de 400 refrigeradores, dos quais 40 possuem compressores com defeito. Se X for o número de compressores defeituosos entre 15 refrigeradores selecionados aleatoriamente, descreva uma forma menos cansativa de calcular (ao menos aproximadamente) $P(X \leq 5)$ em vez de usar a fmp hipergeométrica.

66. Um instrutor que lecionou estatística para engenheiros para duas turmas no semestre passado, a primeira com 20 alunos e a segunda com 30, decidiu pedir aos alunos um projeto semestral. Após a entrega de todos os projetos, o instrutor os organizou aleatoriamente antes de corrigi-los. Considere os primeiros 15 projetos a serem corrigidos.

a. Qual é a probabilidade de exatamente 10 projetos serem da segunda turma?

b. Qual é a probabilidade de pelo menos 10 projetos serem da segunda turma?

c. Qual é a probabilidade de ao menos 10 projetos serem da mesma turma?

d. Qual é o valor médio e o desvio padrão do número de projetos entre esses 15 que pertencem à segunda turma?

e. Qual é o valor médio e o desvio padrão do número de projetos que não estão entre os 15 primeiros que pertencem à segunda turma?

67. Um geólogo coletou 10 amostras de rocha basáltica e 10 de granito. Instruiu o geólogo assistente de laboratório para selecionar aleatoriamente 15 amostras para análise.

a. Qual é o número de amostras de granito selecionadas para análise?

b. Qual é a probabilidade de todas as amostras de um dos dois tipos de rocha serem selecionadas para análise?

c. Qual é a probabilidade de o número de amostras de granito selecionadas para análise estarem dentro de 1 desvio padrão do valor médio?

68. Um diretor de pessoal que entrevistará 11 engenheiros seniores para quatro cargos marcou seis entrevistas para o primeiro dia e cinco para o segundo. Assuma que os candidatos serão entrevistados em ordem aleatória.

a. Qual é a probabilidade de x dos quatro melhores candidatos serem entrevistados no primeiro dia?

b. Dos quatro melhores candidatos, quantos podem ser esperados para entrevistas no primeiro dia?

69. Vinte pares de indivíduos inscritos em um torneio de *bridge* foram designados com os números 1,..., 20. Na primeira parte do torneio, os 20 pares são divididos aleatoriamente em 10 pares leste-oeste e 10 pares norte-sul.

a. Qual é a probabilidade de x dos 10 melhores pares jogarem leste-oeste?

b. Qual é a probabilidade de todos os cinco melhores pares jogarem na mesma direção?

c. Se há $2n$ pares, qual é a fmp de X = número entre os n melhores pares que jogam leste-oeste? Quais são $E(X)$ e $V(X)$?

70. Um alerta de poluição de segundo estágio foi feito em uma determinada área de Los Angeles County, onde há 50 indústrias. Um inspetor visitará 10 indústrias selecionadas aleatoriamente para verificação de violações dos regulamentos.

a. Se 15 empresas realmente estiverem violando ao menos um regulamento, qual é a fmp do número de empresas visitadas pelo inspetor que estão violando ao menos um regulamento?

b. Se houver 500 empresas na área, das quais 150 com violação, faça a aproximação da fmp da parte (a) para uma fmp mais simples.

c. Para X = número entre as 10 empresas visitadas que estão violando os regulamentos, calcule $E(X)$ e $V(X)$ para a fmp exata e aproximada na parte (b).

71. Suponha que $p = P(\text{nascimento de menino}) = 0{,}5$. Um casal quer ter exatamente duas meninas e terá filhos até essa condição ser satisfeita.

a. Qual é a probabilidade de a família ter x filhos homens?

b. Qual é a probabilidade de a família ter quatro filhos?

c. Qual é a probabilidade de a família ter no máximo quatro filhos?

d. Quantos filhos homens espera-se que essa família tenha? Quantos filhos espera-se que essa família tenha?

72. Uma família decide ter filhos até ter três do mesmo sexo. Assumindo $P(H) = P(M) = 0{,}5$, qual é a fmp de X = número de filhos na família?

73. Três irmãos e suas esposas decidem ter filhos até que cada família tenha duas meninas. Qual é a fmp de X = número total de filhos homens nascidos nas famílias? Qual é $E(X)$ e como pode ser comparada ao número esperado de filhos nascidos em cada família?

74. O indivíduo A tem um dado vermelho e B tem um verde (ambos justos). Se cada um lançar o dado até obterem cinco "duplas" (1-1, ..., 6-6), qual será a fmp de X = número total de vezes que um dado é lançado? Quais são $E(X)$ e $V(X)$?

3.6 Distribuição de probabilidade de Poisson

As distribuições binomial, hipergeométrica e binomial negativa foram deduzidas a partir de um experimento consistindo de tentativas ou retiradas e na aplicação das leis da probabilidade aos diversos resultados do experimento. Não há um experimento simples que sirva de base para a distribuição de Poisson, apesar de descrevermos simplificadamente como ela pode ser obtida por meio de certas operações limitantes.

DEFINIÇÃO Uma variável aleatória X tem **distribuição de Poisson** com parâmetro λ ($\lambda > 0$) se a fmp de X for

$$p(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots$$

Geralmente, o valor de λ é uma taxa por unidade de tempo ou por unidade de área. A letra e em $p(x; \lambda)$ é a base dos logaritmos naturais. Seu valor numérico é aproximadamente 2,71828. Como λ é positivo, $p(x; \lambda) > 0$ para todos os valores possíveis de x. O fato de $\sum_{x=0}^{\infty} p(x; \lambda) = 1$ é uma conseqüência da expansão de e^{λ} da série infinita de Maclaurin, que aparece na maioria dos livros-texto de cálculo:

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \tag{3.19}$$

Se os dois termos extremos da Expressão (3.19) forem multiplicados por $e^{-\lambda}$ e depois $e^{-\lambda}$ for colocado dentro da somatória, o resultado será

$$1 = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!}$$

o que mostra que $p(x; \lambda)$ satisfaz a segunda condição necessária para definir uma fmp.

Exemplo 3.38

Seja X o número de certo tipo de animais capturados em uma armadilha durante certo período de tempo. Suponha que X tenha uma distribuição de Poisson com $\lambda = 4{,}5$, de forma que, em média, cada armadilha contém 4,5 animais. [O artigo "Dispersal Dynamics of the Bivalve *Gemma Gemma* in a Patchy Environment (*Ecological Monographs,* 1995, p. 1-20) sugere esse modelo. O bivalvo *Gemma gemma* é um pequeno molusco]. A probabilidade de uma armadilha conter exatamente cinco animais é

$$P(X = 5) = \frac{e^{-4,5}(4{,}5)^5}{5!} = 0{,}1708$$

A probabilidade de uma armadilha conter no máximo cinco animais é

$$P(X \leq 5) = \sum_{x=0}^{5} \frac{e^{-4,5}(4{,}5)^x}{x!} = e^{-4,5}\left[1 + 4{,}5 + \frac{(4{,}5)^2}{2!} + \cdots + \frac{(4{,}5)^5}{5!}\right] = 0{,}7029 \quad \blacksquare$$

A distribuição de Poisson como um limite

A base lógica para o uso da distribuição de Poisson em muitas situações é fornecida pela proposição a seguir.

PROPOSIÇÃO	Suponha que na fmp binomial $b(x; n, p)$, tenhamos $n \to \infty$ e $p \to 0$ de tal forma que np se aproxime de um valor $\lambda > 0$. Então $b(x; n, p) \to p(x; \lambda)$.

De acordo com essa proposição, *em qualquer experimento binomial em que n é grande e p é pequeno,* $b(x; n, p) \approx p(x; \lambda)$, onde $\lambda = np$. Como regra prática, tal aproximação pode ser aplicada se $n \geq 100$, $p \leq 0{,}01$ e $np \leq 20$.

Exemplo 3.39

Se uma editora de livros não-técnicos se esforça para garantir que seus livros não possuem erros tipográficos, de forma que a probabilidade de uma página conter um erro desse tipo é de 0,005 e os erros são independentes de página para página, qual é a probabilidade de um de seus romances de 400 páginas conter exatamente uma página com erros? No máximo três páginas com erros?

Com S representando uma página com ao menos um erro e F uma página sem erros, o número X de páginas que contém ao menos um erro é uma va binomial com $n = 400$ e $p = 0{,}005$, de forma que $np = 2$. Queremos

$$P(X = 1) = b(1; 400, 0{,}005) \approx p(1; 2) = \frac{e^{-2}(2)^1}{1!} = 0{,}271$$

De forma similar,

$$P(X \leq 3) \approx \sum_{x=0}^{3} p(x; 2) = \sum_{x=0}^{3} e^{-2} \frac{2^x}{x!} = 0{,}135 + 0{,}271 + 0{,}271 + 0{,}180$$
$$= 0{,}857$$ ■

A Tabela A.2 do Apêndice mostra a FDA $F(x; \lambda)$ para $\lambda = 0{,}1; 0{,}2;..., 1, 2;..., 10; 15$ e 20. Por exemplo: se $\lambda = 2$, então $P(X \leq 3) = F(3; 2) = 0{,}857$, como no Exemplo 3.39, enquanto $P(X = 3) = F(3; 2) - F(2; 2) = 0{,}180$. De forma alternativa, muitos programas estatísticos gerarам $p(x; \lambda)$ e $F(x; \lambda)$ quando solicitado.

Média e variância de X

Como $b(x; n, p) \to p(x; \lambda)$ quando $n \to \infty$, $p \to 0$, $np \to \lambda$, a média e a variância de uma variável binomial tendem para as de uma variável de Poisson. Os limites são $np \to \lambda$ e $np(1 - p) \to \lambda$.

PROPOSIÇÃO Se X tiver distribuição de Poisson com parâmetro λ, então $E(X) = V(X) = \lambda$.

Esses resultados também podem ser deduzidos diretamente das definições de média e variância.

Exemplo 3.40 (Continuação do Exemplo 3.38)

Tanto o número de animais capturados como a variância dessa quantidade são iguais a 4,5 e $\sigma_X = \sqrt{\lambda} = \sqrt{4{,}5} = 2{,}12$. ■

Processo de Poisson

Uma aplicação muito importante da distribuição de Poisson surge juntamente com a ocorrência de eventos de um tipo particular no decorrer do tempo. Como exemplo, suponha que, iniciando em um instante denominado $t = 0$, estejamos interessados em contar o número de pulsos radioativos registrados em um contador Geiger. Façamos as seguintes hipóteses sobre a maneira como os pulsos ocorrem:

1. Existe um parâmetro $\alpha > 0$ tal que, para qualquer intervalo de tempo curto Δt, a probabilidade de ser recebido exatamente um pulso é $\alpha \cdot \Delta t + o(\Delta t)$.[3]
2. A probabilidade de ser recebido mais de um pulso durante Δt é $o(\Delta t)$ [que, juntamente com a Hipótese 1 implica que a probabilidade de nenhum pulso durante Δt é $1 - \alpha \cdot \Delta t - o(\Delta t)$].
3. O número de pulsos recebidos durante o intervalo de tempo Δt é independente da quantidade recebida antes do intervalo de tempo.

Informalmente, a Hipótese 1 diz que, para um intervalo de tempo curto, a probabilidade de se receber um único pulso é aproximadamente proporcional à duração do intervalo de tempo, onde α é a constante de proporcionalidade. Seja agora $P_k(t)$ a probabilidade de serem recebidos k pulsos pelo contador durante um intervalo de tempo de duração t.

PROPOSIÇÃO $P_k(t) = e^{-\alpha t} \cdot (\alpha t)^k/k!$, de forma que o número de pulsos durante um intervalo de tempo t é uma va de Poisson com parâmetro $\lambda = \alpha t$. O número esperado de pulsos durante qualquer um de tais intervalos de tempo será então αt, de forma que o número esperado durante uma unidade de intervalo de tempo será α.

[3] Uma quantidade é $o(\Delta t)$ (leia "o minúsculo de delta t") se, à medida que Δt se aproxima de 0, $o(\Delta t)/\Delta t$ *também o faz*. Isto é, $o(\Delta t)$ é ainda mais desprezível que o próprio Δt. A quantidade $(\Delta t)^2$ possui essa propriedade, mas $\text{sen}(\Delta t)$ não.

Exemplo 3.41

Suponha que os pulsos cheguem no contador em uma taxa média de seis por minuto, de modo que $\alpha = 6$. Para determinar a probabilidade de pelo menos um pulso ser recebido em um intervalo de meio minuto, observe que o número de pulsos em tal intervalo tem distribuição de Poisson com parâmetro $\alpha t = 6(0,5) = 3$ (0,5 min. é usado porque α é expresso como taxa por minuto). Então, sendo X = número de pulsos recebidos no intervalo de 30 segundos,

$$P(1 \leq X) = 1 - P(X = 0) = 1 - \frac{e^{-3}(3)^0}{0!} = 0{,}950 \quad \blacksquare$$

Se nas Hipóteses 1-3 substituirmos "pulso" por "evento", então o número de eventos que ocorrem durante um intervalo de tempo fixo t tem distribuição de Poisson com parâmetro αt. Qualquer processo que tenha essa distribuição é denominado **processo de Poisson** e α é denominada *taxa do processo*. Outros exemplos de situações que originam um processo de Poisson incluem monitoramento do *estado* de um sistema de computadores ao longo do tempo, com as quebras constituindo os eventos de interesse; registro do número de acidentes em uma instalação industrial ao longo do tempo, tempo de resposta em uma central telefônica; e observação da quantidade de chuveiros de raios cósmicos em um determinado observatório ao longo do tempo.

Em vez de observar eventos ao longo do tempo, considere apenas observar eventos de um determinado tipo que ocorrem em uma região bi ou tridimensional. Por exemplo: podemos selecionar uma determinada região R de uma floresta em um mapa ir para aquela região e contar o número de árvores. Cada árvore representa um evento que ocorre em um ponto certo no espaço. Com hipóteses similares a 1-3, podemos demonstrar que o número de eventos que ocorrem em uma região R tem distribuição de Poisson com parâmetro $\alpha \cdot a(R)$, onde $a(R)$ é a área de R. A quantidade α é o número esperado de eventos por unidade de área ou de volume.

Exercícios | Seção 3.6 (75–89)

75. Seja X o número de falhas na superfície de uma caldeira de um determinado tipo selecionada aleatoriamente, com distribuição de Poisson de parâmetro $\lambda = 5$. Use a Tabela A.2 do Apêndice para calcular as probabilidades a seguir:
a. $P(X \leq 8)$ **b.** $P(X = 8)$ **c.** $P(9 \leq X)$
d. $P(5 \leq X \leq 8)$ **e.** $P(5 < X < 8)$

76. Suponha que o número X de tornados observados em uma determinada região durante o período de um ano tenha uma distribuição de Poisson com $\lambda = 8$.
a. Calcule $P(X \leq 5)$.
b. Calcule $P(6 \leq X \leq 9)$.
c. Calcule $P(10 \leq X)$.
d. Qual é a probabilidade de o número observado de tornados exceder o número esperado por mais de 1 desvio padrão?

77. Suponha que o número de motoristas que trafega entre certa origem e destino durante um período de tempo determinado tenha uma distribuição de Poisson com parâmetro $\lambda = 20$ (sugerido no artigo "Dynamic Ride Sharing: Theory and Practice," J. *of Transp. Engr.*, 1997, p. 308-312). Qual é a probabilidade de o número de motoristas ser:
a. No máximo 10?
b. Exceder 20?
c. Estar entre 10 e 20, inclusive? Estar estritamente entre 10 e 20?
d. Estar dentro de dois desvios padrão do valor médio?

78. Considere gravar algo em um disco rígido e enviá-lo para um certificador contar o número de pulsos ausentes. Suponha que esse número X tenha uma distribuição de Poisson com parâmetro $\lambda = 0{,}2$. (Sugerido em "Average Sample Number for SemiCurtailed Sampling Using the Poisson Distribution," *J. Quality Technology*, 1983, p. 126-129).
a. Qual é a probabilidade de um disco ter exatamente um pulso ausente?
b. Qual é a probabilidade de um disco ter dois pulsos ausentes?
c. Se dois discos forem selecionados de forma independente, qual é a probabilidade de nenhum conter pulsos ausentes?

79. Um artigo no *Los Angeles Times* (3 de dezembro de 1993) relata que 1 em 200 pessoas possui o gene recessivo que causa câncer de cólon hereditário. Em uma amostra de 1000 indivíduos, qual é a distribuição aproximada do número dos que possuem o gene? Use essa distribuição para calcular a probabilidade aproximada de:
a. Entre 5 e 8 (inclusive) possuírem o gene.
b. Ao menos 8 possuírem o gene.

80. Suponha que apenas 0,10% de todos os computadores de certo tipo apresentem falhas de CPU durante o período de garantia. Considere uma amostra de 10.000 computadores.

a. Qual é o valor esperado e o desvio padrão do número de computadores da amostra que apresentam defeito?

b. Qual é a probabilidade (aproximada) de mais de 10 computadores da amostra apresentarem defeito?

c. Qual é a probabilidade (aproximada) de nenhum computador da amostra apresentar defeito?

81. Suponha que pequenas aeronaves pousem em um aeroporto, de acordo com um processo de Poisson, com a taxa $\alpha = 8$ por hora, de forma que o número de pousos durante um período de tempo de t horas é uma va de Poisson com parâmetro $\lambda = 8t$.

a. Quais são a probabilidade de exatamente seis aeronaves pequenas chegarem durante um período de uma hora? Ao menos seis? Ao menos 10?

b. Qual é o valor esperado e o desvio padrão do número de pequenas aeronaves que chegam durante um período de 90 minutos?

c. Qual é a probabilidade de ao menos 20 aeronaves pequenas chegarem durante um período de $2\frac{1}{2}$ horas? De no máximo 10 chegarem nesse período?

82. O número de pessoas que chegam para tratamento em um pronto-socorro pode ser modelado por um processo de Poisson com taxa de cinco por hora.

a. Qual é a probabilidade de exatamente quatro pessoas chegarem em certa hora?

b. Qual é a probabilidade de ao menos quatro pessoas chegarem em certa hora?

c. Quantas pessoas você espera que cheguem em um período de 45 minutos?

83. O número de solicitações de assistência recebido por um serviço de guincho é um processo de Poisson com taxa $\alpha = 4$ por hora.

a. Calcule a probabilidade de exatamente dez solicitações chegarem em um certo período de 2 horas.

b. Se os operadores do serviço de guincho tirarem 30 minutos para almoço, qual é a probabilidade de não perderem nenhum chamado de assistência?

c. Quantas ligações você espera que ocorram durante o almoço?

84. Em testes de placas de circuitos, a probabilidade de falha em um diodo é de 0,01. Suponha que uma placa de circuito contenha 200 diodos.

a. Quantos diodos espera-se que apresentem falhas e qual é o desvio padrão desse valor?

b. Qual é a probabilidade (aproximada) de ao menos quatro diodos apresentarem falha em uma placa selecionada aleatoriamente?

c. Se cinco placas forem enviadas a um determinado cliente, qual é a probabilidade de ao menos quatro funcionarem corretamente? (Uma placa só funciona corretamente se todos os seus diodos funcionarem.)

85. O artigo "Reliability-Based Service-Life Assessment of Aging Concrete Structures" (*J. Structural Engr.*, 1993, p. 1600-1621) sugere que um processo de Poisson pode ser usado para representar a ocorrência de cargas estruturais no tempo. Suponha que o tempo médio entre as ocorrências de cargas seja 0,5 ano.

a. Quantas cargas podem ser esperadas durante um período de dois anos?

b. Qual é a probabilidade de mais de cinco cargas ocorrerem durante um período de dois anos?

c. Quanto tempo deve ter um período para que a probabilidade de não ocorrerem cargas seja no máximo 0,1?

86. Seja X uma distribuição de Poisson com parâmetro λ. Demonstre que $E(X) = \lambda$ diretamente da definição de valor esperado. (*Sugestão*: O primeiro termo da soma é igual a 0, então x pode ser cancelado. Fatore λ e mostre que o restante soma 1.)

87. Suponha que as árvores sejam distribuídas em uma floresta de acordo com um processo bidimensional de Poisson com parâmetro α. O número esperado de árvores por acre é igual a 80.

a. Qual é a probabilidade de que em um quarto de acre haja no máximo 16 árvores?

b. Se a floresta cobrir 85.000 acres, qual será o número esperado de árvores na floresta?

c. Suponha que você selecione um ponto na floresta e construa um círculo de raio 0,1 milha. Seja X = número de árvores na região circular. Qual é a fmp de X? (*Sugestão*: 1 milha quadrada = 640 acres.)

88. Automóveis chegam em uma estação de inspeção de equipamentos veiculares de acordo com um processo de Poisson com taxa $\alpha = 10$ por hora. Suponha que, com probabilidade 0,5, um veículo chegue sem violações de equipamento.

a. Qual é a probabilidade de exatamente 10 veículos chegarem durante uma hora e nenhum ter violações?

b. Para qualquer $y \geq 10$ fixo, qual é a probabilidade de y veículos chegarem em uma hora, todos sem violações?

c. Qual é a probabilidade de 10 carros sem violações chegarem na próxima hora? [*Sugestão*: Some as probabilidades na parte (b) entre $y = 10$ e ∞.]

89. a. Em um processo de Poisson, o que deve acontecer no intervalo de tempo $(0, t)$ e no intervalo $(t, t + \Delta t)$ para que nenhum evento ocorra em todo o intervalo $(0, t + \Delta t)$? Use esta Hipótese e as Hipóteses 1-3 para escrever a relação entre $P_0(t + \Delta t)$ e $P_0(t)$.

b. Use o resultado da parte (a) para escrever uma expressão para a diferença $P_0(t + \Delta t) - P_0(t)$. Divida então por Δt e considere que $\Delta t \to 0$ para obter uma

equação envolvendo $(d/dt)P_0(t)$, derivada de $P_0(t)$ em relação a t.

c. Verifique se $P_0(t) = e^{-\alpha t}$ satisfaz à equação da parte (b).

d. Pode-se demonstrar de forma similar às partes (a) e (b) que as $P_k(t)$ devem satisfazer ao sistema de equações diferenciais.

$$\frac{d}{dt}P_k(t) = \alpha P_{k-1}(t) - \alpha P_k(t)$$

$$k = 1, 2, 3, \ldots$$

Verifique se $P_k(t) = e^{-\alpha t}(\alpha t)^k/k!$ satisfaz ao sistema. (Na verdade, esta é a única solução.)

Exercícios suplementares (90–116)

90. Considere um baralho formado por sete cartas, marcadas 1, 2, ... , 7. Três dessas cartas são selecionadas aleatoriamente. Defina uma va W como W = soma dos números resultantes e calcule a fmp de W. Depois, calcule μ e σ^2. [*Sugestão*: Considere os resultados como não-ordenados, de forma que (1, 3, 7) e (3, 1, 7) não sejam considerados resultados diferentes. Então haverá 35 resultados que podem ser relacionados. (Esse tipo de va, na verdade, surge juntamente com o teste de hipótese denominado teste de soma de postos de Wilcoxon, em que há uma amostra x e uma amostra y e W é a soma dos postos dos x na amostra combinada.)]

91. Após embaralhar um baralho de 52 cartas, um crupiê dá cinco cartas. Seja X = número de naipes representados na mão de cinco cartas.

a. Mostre que a fmp de X é

x	1	2	3	4
$p(x)$	0,002	0,146	0,588	0,264

[*Sugestão*: $p(1) = 4P$(todos de espadas), $p(2) = 6P$(apenas espadas e copas com no mínimo um de cada) e $p(4) = 4P$(2 espadas $\cap$ um de cada outro naipe).]

b. Calcule μ, σ^2, e σ.

92. A va binomial negativa X foi definida como o número de F's antes do r-ésimo S. Seja Y = número de tentativas necessárias para obtenção do r-ésimo S. Da mesma forma que a fmp de X foi deduzida, deduza a fmp de Y.

93. De todos os clientes que compram portões automáticos de garagem, 75% compram um modelo acionado por corrente. Seja X = número entre os próximos 15 compradores que selecionam o modelo acionado por corrente.

a. Qual é a fmp de X?

b. Calcule $P(X > 10)$.

c. Calcule $P(6 \le X \le 10)$.

d. Calcule μ e σ^2.

e. Se a loja atualmente tiver em estoque 10 modelos acionados por corrente e oito acionados por eixo, qual é a probabilidade de os pedidos desses 15 clientes serem todos atendidos com o estoque existente?

94. Um amigo recentemente planejou uma viagem para um acampamento. Ele tinha duas lanternas, uma que exigia uma pilha de 6V e outra que usava duas pilhas tamanho D. Ele já havia colocado na mala duas pilhas de 6V e quatro tamanho D. Suponha que a probabilidade de uma determinada pilha funcionar seja p e que as pilhas funcionam ou não, independentemente uma da outra. Nosso amigo quer levar apenas uma lanterna. Para que valores de p ele deve levar a lanterna de 6V?

95. Um *sistema k-de-n* funciona se, e somente se, pelo menos k dos n componentes individuais do sistema funcionarem. Se os componentes individuais funcionam independentemente um do outro, cada um com probabilidade 0,9, qual é a probabilidade de um sistema 3-de-5 funcionar?

96. Um fabricante de pilhas de lanterna deseja controlar a qualidade de seus produtos rejeitando os lotes em que a proporção de pilhas com voltagem inaceitável seja muito alta. Assim, em cada lote grande (10.000 pilhas), 25 serão selecionadas e testadas. Se ao menos cinco delas gerarem voltagens inaceitáveis, todo o lote será rejeitado. Qual é a probabilidade de um lote ser rejeitado se

a. 5% das pilhas do lote tiverem voltagens inaceitáveis?

b. 10% das pilhas do lote tiverem voltagens inaceitáveis?

c. 20% das pilhas do lote tiverem voltagens inaceitáveis?

d. O que aconteceria às probabilidades nas partes (a)–(c) se o número crítico de rejeição fosse aumentado de 5 para 6?

97. Das pessoas que passam por um detector de metal de um aeroporto, 0,5% o ativam. Seja X = número de pessoas que ativam o detector entre um grupo de 500 selecionado aleatoriamente.

a. Qual é a fmp (aproximada) de X?

b. Calcule $P(X = 5)$.

c. Calcule $P(5 \le X)$.

98. Uma empresa de consultoria educacional está tentando decidir se alunos de ensino médio que nunca usaram uma calculadora podem resolver um tipo de problema mais facilmente com uma calculadora que usa lógica polonesa reversa ou com uma que não a usa. Uma amostra de 25 alunos é selecionada e eles podem praticar em ambas as calculadoras. Então, cada aluno é solicitado a resolver um problema com a

calculadora polonesa reversa e um similar com a outra. Seja $p = P(S)$, onde S indica que o aluno resolveu o problema mais rapidamente com a lógica polonesa reversa do que com a outra e seja X = número de $S's$.

a. Se $p = 0,5$, qual é $P(7 \leq X \leq 18)$?

b. Se $p = 0,8$, qual é $P(7 \leq X \leq 18)$?

c. Se a Hipótese $p = 0,5$ for rejeitada quando $X \leq 7$ ou $X \geq 18$, qual é a probabilidade de rejeitar a hipótese quando ela estiver correta?

d. Se a decisão de rejeitar a hipótese $p = 0,5$ for como na parte (c), qual é a probabilidade de ela não ser rejeitada quando $p = 0,6$? E quando $p = 0,8$?

e. Que regra de decisão você escolheria para rejeitar a hipótese $p = 0,5$ se você quisesse que a probabilidade da parte (c) fosse no máximo 0,01?

99. Considere uma doença cuja presença pode ser identificada por um teste sangüíneo. Seja p a probabilidade de um indivíduo selecionado aleatoriamente ter a doença. Suponha que n indivíduos sejam selecionados independentemente para teste. Uma forma de proceder é fazer um teste separado para cada uma das n amostras sangüíneas. Uma abordagem potencialmente mais econômica — teste em grupo — foi apresentada durante a Segunda Guerra Mundial para identificar homens com sífilis entre recrutas do exército. Primeiro, separe parte de cada amostra de sangue, misture-as e faça um único teste. Se ninguém tiver a doença, o resultado será negativo e apenas um teste terá sido necessário. Se ao menos um indivíduo tiver a doença, o teste da amostra misturada resultará positivo e nesse caso serão processados os n testes individuais. Se $p = 0,1$ e $n = 3$, qual será o número esperado de testes usando esse procedimento? Qual será o número esperado quando $n = 5$? [O artigo "Random Multiple Access Communication and Group Testing" (*IEEE Trans. on Commun.*, 1984, p. 769-774) aplicou essas idéias a um sistema de comunicação em que a dicotomia era usuário ativo/ocioso em vez de doente/não-doente.]

100. Seja p_1 a probabilidade de qualquer símbolo de certo código ser transmitido com erro em um sistema de comunicação. Assuma que os erros ocorram em símbolos diferentes, independentemente um do outro. Suponha também que com probabilidade p_2, o símbolo errado seja corrigido no recebimento. Seja X o número de símbolos corrigidos em um bloco de mensagem consistindo de n símbolos (após o final do processo de correção). Qual é a distribuição de probabilidade de X?

101. O comprador de um gerador exige c partidas consecutivas antes de acertar a unidade. Assuma que os resultados das partidas individuais sejam independentes um do outro. Seja p a probabilidade de certa partida ter sucesso. A variável aleatória de interesse é X = número de partidas que devem ser processadas antes da aceitação. Forneça a fmp de X para o caso $c = 2$. Se $p = 0,9$, qual será $P(X \leq 8)$? [*Sugestão*: para $x \geq 5$, expresse $p(x)$ "de forma recursiva" em termos da fmp calculada com os valores menores $x - 3, x - 4, \ldots, 2$.] (Esse problema foi sugerido pelo artigo "Evaluation of a Start-Up Demonstration Test," *J. Quality Technology*, 1983, p. 103-106.)

102. Um plano de fidelidade de viagens de executivos foi desenvolvido por uma companhia aérea sob a premissa de que 10% de seus clientes atuais se qualificariam para participar.

a. Admitindo a validade dessa premissa, entre 25 clientes atuais selecionados aleatoriamente, qual é a probabilidade de entre 2 e 6 (inclusive) se qualificarem para o programa?

b. Admitindo novamente a validade da premissa, qual é o número esperado de clientes que se qualificam e o desvio padrão do número dos que se qualificam em uma amostra aleatória de 100 clientes?

c. Seja X o número em uma amostra de 25 clientes atuais que se qualificam para o programa. Considere rejeitar a premissa da empresa na hipótese de $p > 0,10$ se $x \geq 7$. Qual é a probabilidade de a premissa da empresa ser rejeitada quando for válida?

d. Refira-se à regra de decisão apresentada na parte (c). Qual é a probabilidade de a premissa da empresa não ser rejeitada mesmo que $p = 0,20$ (ou seja, 20% se qualificam)?

103. Quarenta por cento das sementes de milho possuem espigas simples e as outras 60%, espigas duplas. Uma semente com espiga simples produzirá uma planta com espigas simples em 29% das vezes, enquanto uma semente com espiga dupla produzirá uma planta com espigas simples em 26% das vezes. Considere a seleção aleatória de 10 sementes.

a. Qual é a probabilidade de exatamente cinco dessas sementes terem espiga única e produzirem uma planta com espigas únicas?

b. Qual é a probabilidade de exatamente cinco espigas possuírem espigas únicas? Qual é a probabilidade de no máximo cinco plantas terem espigas únicas?

104. Um julgamento não resultou em veredicto, porque oito membros do júri eram favoráveis ao veredicto de culpado e os outros quatro eram favoráveis ao de absolvição. Se os jurados saírem da sala aleatoriamente e cada um dos quatro primeiros for interpelado por um repórter buscando uma entrevista, qual será a fmp de X = número de jurados favoráveis à absolvição, entre os entrevistados? Quantos dos favoráveis à absolvição, você espera que sejam entrevistados?

105. Um serviço de reservas emprega cinco operadores de informações que recebem solicitações de informações independentemente um do outro de acordo com um processo de Poisson com taxa $\alpha = 2$ por minuto.

a. Qual é a probabilidade de que, durante certo período de 1 minuto, o primeiro operador não receba solicitações?

b. Qual é a probabilidade de que, durante certo período de um minuto, quatro dos cinco operadores não recebam solicitações?

c. Escreva uma expressão para a probabilidade de em um dado período de um minuto, todos os operadores recebam exatamente o mesmo número de solicitações.

106. Gafanhotos distribuem-se aleatoriamente em um vasto campo, de acordo com uma distribuição de Poisson, com parâmetro $\alpha = 2$ por jarda quadrada. Qual é o tamanho do raio R de uma região amostral circular para que a probabilidade de se encontrar pelo menos um gafanhoto na região seja de 0,99?

107. Uma banca de jornal faz o pedido de cinco cópias de uma edição de certa revista de fotografia. Seja X = número de indivíduos que chegam para comprar a revista. Se X tiver distribuição de Poisson com parâmetro $\lambda = 4$, qual será o número esperado de cópias vendidas?

108. Os indivíduos A e B começam a jogar uma seqüência de partidas de xadrez. Seja $S = \{$A ganha a partida$\}$ e suponha que os resultados de partidas sucessivas sejam independentes com $P(S) = p$ e $P(F) = 1 - p$ (nunca há empate). Eles jogarão até que um ganhe 10 partidas. Seja X = número de partidas jogadas (com valores possíveis 10, 11,..., 19).

a. Para x = 10, 11,..., 19, obtenha a expressão de $p(x) = P(X = x)$.

b. Se puder haver empates, com $p = P(S)$, $q = P(F)$, $1 - p - q = P$ (empate), quais serão os valores possíveis de X? Qual é $P(20 \leq X)$? [*Sugestão*: $P(20 \leq X) = 1 - P(X < 20)$.]

109. Certo teste de presença de uma certa doença tem probabilidade 0,20 de fornecer um resultado positivo falso (indicando que o indivíduo tem a doença quando não for o caso) e probabilidade 0,10 de fornecer um resultado negativo falso. Suponha que 10 indivíduos sejam testados, cinco dos quais têm a doença e cinco não. Seja X = número de resultados positivos.

a. X tem distribuição binomial? Explique seu raciocínio.

b. Qual é a probabilidade de exatamente três dos 10 resultados serem positivos?

110. A fmp geral da binomial negativa é dada por

$$nb(x; r, p) = k(r, x) \cdot p^r(1 - p)^x$$
$$x = 5\ 0, 1, 2, \ldots$$

Seja X o número de certo tipo de planta encontrada em uma determinada região, calcule a distribuição para $p = 0{,}3$ e $r = 2{,}5$. Qual é $P(X = 4)$? Qual é a probabilidade de ao menos uma planta ser encontrada?

111. Defina uma função $p(x; \lambda, \mu)$ por

$$p(x; \lambda, \mu) = \begin{cases} \frac{1}{2}e^{-\lambda}\frac{\lambda^x}{x!} + \frac{1}{2}e^{-\mu}\frac{\mu^x}{x!} & x = 0, 1, 2, \ldots \\ 0 & \text{caso contrário} \end{cases}$$

a. Mostre que $p(x; \lambda, \mu)$ satisfaz às duas condições necessárias para especificação de uma fmp. [*Nota*: Se uma empresa emprega dois digitadores, um dos quais comete erros tipográficos na taxa de λ por página e o outro na taxa de μ por página e cada um deles faz metade do trabalho de digitação da empresa, então $p(x; \lambda, \mu)$ é a fmp de X = número de erros em uma página selecionada aleatoriamente.]

b. Se o primeiro digitador (taxa λ) digita 60% de todas as páginas, qual é a fmp de X da parte (a)?

c. Qual é $E(X)$ de $p(x; \lambda, \mu)$ dado por essa expressão?

d. Qual é o valor de σ^2 de $p(x; \lambda, \mu)$ obtido por aquela expressão?

112. A *moda* de uma variável aleatória discreta X com fmp $p(x)$ é o valor x^* para o qual $p(x)$ é o maior (o valor mais provável de x).

a. Seja $X \sim \text{Bin}(n, p)$. Considerando a razão $b(x + 1; n, p)/b(x; n, p)$, demonstre que $b(x; n, p)$ aumenta com x, desde que $x < np - (1 - p)$. Conclua que a moda x^* é o inteiro que satisfaz $(n + 1)p - 1 \leq x^* \leq (n + 1)p$.

b. Demonstre que, se X tem distribuição de Poisson com parâmetro λ, a moda é o maior inteiro menor que λ. Se λ for um inteiro, mostre que $\lambda - 1$ e λ são modas.

113. O dispositivo de armazenagem de um computador possui 10 faixas concêntricas numeradas 1, 2, ..., 10 da mais externa para a mais interna e um único braço de acesso. Seja p_i = a probabilidade de certa solicitação de dados levar o braço para a faixa i (i = 1,...,10). Assuma que as faixas acessadas em buscas sucessivas são independentes. Seja X = número de faixas sobre as quais o braço de acesso passa durante duas solicitações sucessivas (excluindo a faixa da qual o braço acaba de sair, de forma que os valores possíveis de X são x = 0, 1,..., 9). Calcule a fmp de X. [*Sugestão*: P(o braço está na faixa i e $X = j$) $= P(X = j \mid$ braço em $i) \cdot p_i$. Após a probabilidade condicional ser expressa em termos de $p_1, \ldots, p_{10}$ pela lei da probabilidade total, a probabilidade desejada é obtida pela somatória dos valores i.]

114. Se X é uma va hipergeométrica, demonstre diretamente pela definição que $E(X) = nM/N$ (considere apenas o caso $n < M$). [*Sugestão*: Fatore nM/N na soma de $E(X)$ e mostre que os termos da soma têm a forma $h(y; n - 1, M - 1, N - 1)$, onde $y = x - 1$.]

115. Use o fato de

$$\sum_{\text{todos } x} (x - \mu)^2 p(x) \geq \sum_{x: |x-\mu| \geq k\sigma} (x - \mu)^2 p(x)$$

para demonstrar a desigualdade de Chebyshev dada no Exercício 43.

116. O processo de Poisson simples da Seção 3.6 é caracterizado por uma taxa constante α em que os eventos ocorrem por unidade de tempo. Uma generalização disso é supor a probabilidade de ocorrência de

exatamente um evento no intervalo $[t, t + \Delta t]$ é $\alpha(t) \cdot \Delta t + o(\Delta t)$. Pode-se, então, demonstrar que o número de eventos que ocorrem durante um intervalo $[t_1, t_2]$ tem uma distribuição de Poisson com parâmetro

$$\lambda = \int_{t_1}^{t_2} \alpha(t)\, dt$$

A ocorrência de eventos ao longo do tempo nessa situação é denominada *processo de Poisson não-homogêneo*. O artigo "Inference Based on Retrospective Ascertainment", *J. Amer Stat. Assoc.*, 1989, p. 360-372, considera a função intensidade

$$\alpha(t) = e^{a+bt}$$

apropriada para eventos que envolvem a transmissão do HIV (vírus da AIDS) via transfusões de sangue. Suponha que $a = 2$ e $b = 0{,}6$ (próximos aos valores sugeridos no artigo), com o tempo em anos.

a. Qual é o número esperado de eventos no intervalo [0, 4]? Em [2, 6]?

b. Qual é a probabilidade de no máximo 15 eventos ocorrerem no intervalo [0, 0,9907]?

Bibliografia

JOHNSON, Norman, KOTZ, Samuel e KEMP, Adrienne, *Discrete Univariate Distributions*. Wiley, Nova York, 1992. Uma enciclopédia de informações sobre distribuições discretas.

OLKIN, Ingram, DERMAN, Cyrus e GLESER, Leon, *Probability Models and Applications* (2. ed.), Macmillan, Nova York, 1994. Contém uma profunda discussão das propriedades gerais de distribuições discretas e contínuas e resultados para distribuições específicas.

ROSS, Sheldon, *Introduction to Probability Models* (7 ed.), Academic Press, Nova York, 2003. Boa fonte de material sobre o processo de Poisson e generalizações, além de ser uma boa introdução a outros tópicos de probabilidade aplicada.

4

Variáveis Aleatórias Contínuas e Distribuições de Probabilidade

Introdução

Como mencionado no início do Capítulo 3, nem todas as variáveis aleatórias são discretas. Neste capítulo, estudamos o segundo tipo geral de variável aleatória que aparece em diversos problemas práticos. As Seções 4.1 e 4.2 apresentam as definições básicas e as propriedades das variáveis aleatórias contínuas e suas distribuições de probabilidades. Na Seção 4.3, estudamos em detalhe a variável aleatória normal e sua distribuição, inquestionavelmente a mais importante e útil da probabilidade e estatística. As seções 4.4 e 4.5 discutem outras distribuições contínuas freqüentemente usadas no trabalho prático. Na Seção 4.6, apresentamos um método de avaliação da consistência dos dados da amostra com a distribuição especificada.

4.1 Variáveis Aleatórias Contínuas e Funções de Densidade de Probabilidade

Uma variável aleatória (va) é discreta se os seus valores possíveis constituírem tanto um conjunto finito como puderem ser relacionados em uma seqüência infinita (uma lista em que haja um primeiro elemento, um segundo etc.). Uma variável aleatória cujo conjunto de valores possíveis consiste de um intervalo completo de números não é discreta.

Variáveis Aleatórias Contínuas

DEFINIÇÃO — Uma variável aleatória X é dita contínua se o seu conjunto de valores possíveis consistir do intervalo completo de todos os valores, isto é, se, para cada $A < B$, qualquer valor x entre A e B for possível.

Exemplo 4.1

Se no estudo de ecologia de um lago fizermos medidas de profundidade em locais selecionados aleatoriamente, então X = profundidade nesse local é uma va contínua. Neste caso, A é a profundidade mínima na região da amostragem e B é a profundidade máxima. ■

Exemplo 4.2

Se um composto químico for selecionado aleatoriamente e determinarmos seu pH X, então, X é uma va contínua porque qualquer valor de PH entre 0 e 14 é possível. Caso se saiba mais sobre o composto selecionado para análise, o conjunto de valores possíveis pode ser um subintervalo de [0, 14], tal como $5{,}5 \leq x \leq 6{,}5$, mas X ainda seria contínua. ■

Se a escala de medida de X puder ser subdividida tanto quanto se desejar, a variável será contínua. Caso contrário, a variável será discreta. Por exemplo: se a variável for altura ou comprimento, poderá ser medida em quilômetros, metros, centímetros e assim por diante, de forma que a variável é contínua. Se, entretanto, X = valor de uma conta de gás mensal selecionada aleatoriamente, a menor unidade de medida será o centavo, fazendo com que qualquer valor de X seja um múltiplo de US$ 0,01 e X seja uma variável discreta.

Alguém pode argumentar que variáveis como altura, peso e temperatura, apesar de serem, em princípio, contínuas, na prática, as limitações dos instrumentos de medida nos restringem a um mundo discreto (embora possa ser bastante subdividido). Entretanto, os modelos contínuos normalmente se aproximam muito bem de situações do mundo real e a matemática contínua (cálculo) freqüentemente se mostra mais fácil de trabalhar do que a matemática de variáveis e distribuições discretas.

Distribuições de Probabilidade de Variáveis Contínuas

Suponha que a variável de interesse X seja a profundidade de um lago em um ponto da superfície escolhido aleatoriamente. Seja M = profundidade máxima (em metros), de modo que qualquer número do intervalo $[0, M]$ seja um valor possível de X. Se "considerarmos X discreta", arredondando a profundidade para o valor mais próximo de metro, então os valores possíveis são inteiros não-negativos menores ou iguais a M. A distribuição discreta resultante das profundidades pode ser ilustrada usando-se um histograma de probabilidade. Se desenharmos o histograma de forma que a área do retângulo acima de qualquer inteiro possível k seja a proporção do lago cuja profundidade seja k (com arredondamento para o metro seguinte), então a área total de todos os retângulos será 1. Um histograma possível é exibido na Figura 4.1(a).

Se a profundidade for medida com arredondamento para o centímetro seguinte e for usado o mesmo eixo de medidas da Figura 4.1(a), cada retângulo no histograma de probabilidades será muito mais estreito, apesar da área total de todos os retângulos permanecer 1. Um histograma possível é ilustrado na Figura 4.1(b). Sua aparência é mais ajustada do que o histograma da Figura 4.1(a). Se continuarmos executando medidas de forma mais e mais precisa, a seqüência resultante terá uma curva cada vez mais ajustada, como a ilustrada na Figura 4.1(c). Como, para cada histograma, a área de todos os retângulos é igual a 1, a área total sob a curva ajustada também é 1. A probabilidade de a profundidade em um ponto selecionado aleatoriamente estar entre a e b é igual à área sob a curva ajustada entre os dois referidos pontos a e b. É exatamente uma curva ajustada do tipo ilustrado na Figura 4.1(c) que especifica uma distribuição de probabilidade contínua.

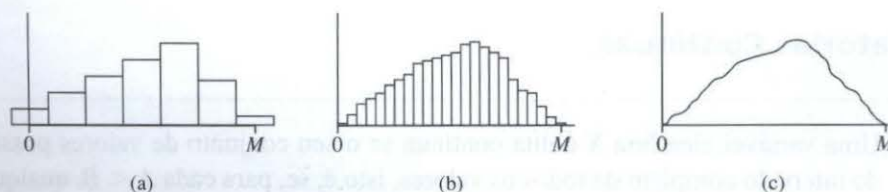


Figura 4.1 (a) Histograma de probabilidade da profundidade medida com arredondamento em metros; (b) histograma de probabilidade da profundidade medida com arredondamento em centímetros; (c) limite de uma seqüência de histogramas discretos

DEFINIÇÃO

Seja X uma va contínua. A **distribuição de probabilidade** ou **função de densidade de probabilidade** (fdp) de X será, então, uma função $f(x)$ tal que, para quaisquer dois números a e b com $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

Isto é, a probabilidade de X ter um determinado valor no intervalo $[a, b]$ é a área contida entre o intervalo e abaixo da curva da função de densidade, conforme ilustrado na Figura 4.2. O gráfico de $f(x)$ normalmente é denominado *curva de densidade*.

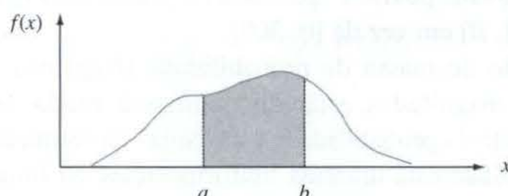


Figura 4.2 $P(a \leq X \leq b)$ = a área abaixo da curva de densidade, entre a e b

Para que $f(x)$ seja uma fdp legítima, deve satisfazer às duas condições a seguir:

1. $f(x) \geq 0$ para todos os x
2. $\int_{-\infty}^{\infty} f(x)\, dx$ = área abaixo do gráfico de $f(x)$
   $= 1$

Exemplo 4.3

A direção de uma imperfeição em relação a uma linha de referência em um objeto circular como um pneu, um rotor de freio ou um volante normalmente apresenta alguma incerteza. Considere a linha de referência que conecta a válvula do pneu até o ponto central e seja X o ângulo medido no sentido horário até o local da imperfeição. Uma fdp possível de X é

$$f(x) = \begin{cases} \dfrac{1}{360} & 0 \leq x < 360 \\ 0 & \text{caso contrário} \end{cases}$$

A fdp está ilustrada na Figura 4.3. Claramente, $f(x) \geq 0$. A área abaixo da curva de densidade é apenas a área de um retângulo: (altura)(base) $= (\frac{1}{360})(360) = 1$. A probabilidade de o ângulo estar entre 90° e 180° é

$$P(90 \leq X \leq 180) = \int_{90}^{180} \frac{1}{360}\, dx = \frac{x}{360}\Big|_{x=90}^{x=180} = \frac{1}{4} = 0{,}25$$

A probabilidade de o ângulo de ocorrência estar dentro de 90° da linha de referência é

$$P(0 \leq X \leq 90) + P(270 \leq X < 360) = 0{,}25 + 0{,}25 = 0{,}50$$

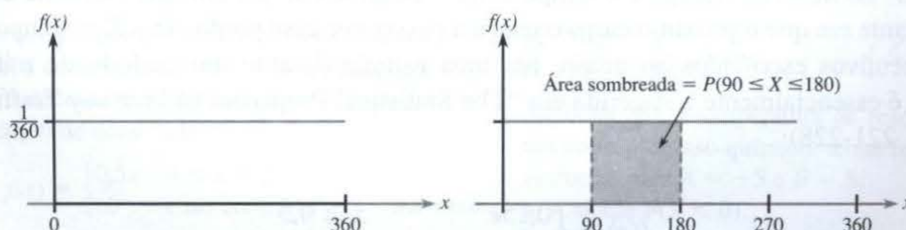


Figura 4.3 A fdp e a probabilidade do Exemplo 4.3

■

Como sempre $0 \leq a \leq b \leq 360$ no Exemplo 4.3, $P(a \leq X \leq b)$ depende apenas da largura $b - a$ do intervalo, X é dita ter distribuição uniforme.

DEFINIÇÃO

Uma va contínua X é dita ter **distribuição uniforme** no intervalo $[A, B]$ se a fdp de X for

$$f(x; A, B) = \begin{cases} \dfrac{1}{B - A} & A \leq x \leq B \\ 0 & \text{caso contrário} \end{cases}$$

O gráfico de qualquer fdp uniforme possui a aparência do gráfico da Figura 4.3, exceto pelo fato de que o intervalo da densidade positiva é $[A, B]$ em vez de $[0, 360]$.

No caso discreto, uma função de massa de probabilidade (fmp) nos diz como as pequenas "falhas" daquele tipo de massa, de diversas magnitudes, estão distribuídas de modo contínuo ao longo do eixo das medidas. No caso contínuo, a densidade da probabilidade está "suja" de forma contínua ao longo do intervalo de valores possíveis. Quando a densidade está mantida uniformemente ao longo do intervalo, resulta uma fdp uniforme, como na Figura 4.3.

Quando X é uma va discreta, é designada uma probabilidade positiva a cada valor possível, mas este não é o caso quando X é contínua.

PROPOSIÇÃO

Se X é uma va contínua, então para qualquer número c, $P(X = c) = 0$. Além disso, para quaisquer dois números a e b com $a < b$,

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) \\ &= P(a \leq X < b) = P(a < X < b) \end{aligned} \tag{4.1}$$

Em palavras, a probabilidade atribuída a qualquer valor específico é zero e a probabilidade de um intervalo não depende da inclusão ou não de seus pontos extremos. Essas propriedades decorrem dos fatos de que a área abaixo do gráfico de $f(x)$ e acima do único valor c é nula, e que a área sob o gráfico acima de um intervalo não é afetada pela exclusão ou inclusão dos pontos extremos do intervalo. Se X for discreta e houver uma massa de probabilidade positiva tanto em $X = a$ como em $X = b$, então as quatro probabilidades da Expressão (4.1) diferirão uma da outra.

O fato de uma distribuição contínua atribuir probabilidade zero para cada valor tem uma analogia física. Considere uma haste circular sólida com área da seção = 1 pol². Posicione a haste ao longo do eixo das medidas e suponha que sua densidade em qualquer ponto x seja dada pelo valor $f(x)$ de uma função de densidade. Então, se a haste for seccionada nos pontos a e b e o segmento removido, a quantidade de massa removida é $\int_a^b f(x)\,dx$. Se a haste for cortada apenas no ponto c, nenhuma massa será removida. A massa é atribuída a segmentos do intervalo da haste, mas não a pontos individuais.

Exemplo 4.4

"Tempo de avanço" no fluxo do tráfego é o tempo entre o instante em que um carro termina de passar por um ponto fixo e o instante em que o próximo carro começa a passar por esse ponto. Seja X = tempo de avanço para dois carros consecutivos escolhidos ao acaso, em uma estrada durante um período de tráfego intenso. A seguinte fdp de X é essencialmente a sugerida em "The Statistical Properties of Freeway Traffic" (*Transp. Research*, vol. 11, p. 221-228):

$$f(x) = \begin{cases} 0{,}15e^{-0{,}15(x-0{,}5)} & x \geq 0{,}5 \\ 0 & \text{caso contrário} \end{cases}$$

O gráfico de $f(x)$ é mostrado na Figura 4.4. Não há densidade associada aos tempos de avanço inferiores a 0,5 e a densidade do avanço decresce exponencialmente rápida à medida que x aumenta a partir de 0,5. Claramente, $f(x) \geq 0$. Para mostrar que $\int_{-\infty}^{\infty} f(x)\,dx = 1$, usamos a expressão do cálculo $\int_a^{\infty} e^{-kx}\,dx = (1/k)e^{-k \cdot a}$. Então

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0,5}^{\infty} 0{,}15e^{-0{,}15(x-0{,}5)}\,dx = 0{,}15e^{0{,}075}\int_{0,5}^{\infty} e^{-0{,}15x}\,dx$$

$$= 0{,}15e^{0{,}075} \cdot \frac{1}{0{,}15}e^{-(0{,}15)(0{,}5)} = 1$$

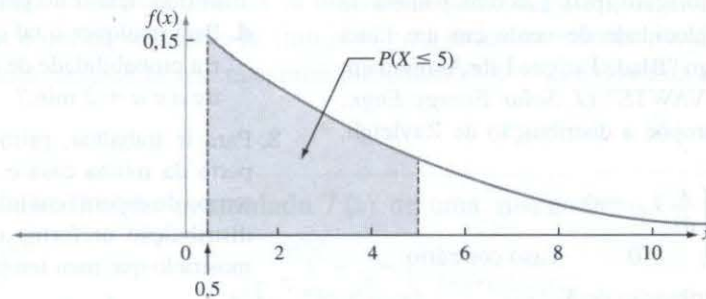


Figura 4.4 A curva de densidade do tempo de avanço do Exemplo 4.4

A probabilidade do tempo de avanço ser no máximo 5 segundos é

$$P(X \leq 5) = \int_{-\infty}^{5} f(x)\,dx = \int_{0,5}^{5} 0{,}15e^{-0{,}15(x-0{,}5)}\,dx$$

$$= 0{,}15e^{0{,}075}\int_{0,5}^{5} e^{-0{,}15x}\,dx = 0{,}15e^{0{,}075} \cdot -\frac{1}{0{,}15}e^{-0{,}15x}\Big|_{x=0,5}^{x=5}$$

$$= e^{0{,}075}(-e^{-0{,}75} + e^{-0{,}075}) = 1{,}078(-0{,}472 + 0{,}928) = 0{,}491$$

$$= P(\text{menos de 5 segundos}) = P(X < 5)$$ ■

Diferentemente das distribuições discretas, tais como a binomial, a hipergeométrica e a binomial negativa, a distribuição de qualquer va contínua dada normalmente não pode ser deduzida por meio de argumentos probabilísticos simples. Pelo contrário, deve-se fazer uma escolha cuidadosa da fdp com base em conhecimentos anteriores e dados disponíveis. Felizmente, há algumas famílias gerais de fdps que funcionam bem em diversas situações experimentais, várias delas discutidas posteriormente, neste capítulo.

Assim como no caso discreto, normalmente é útil imaginar a população de interesse constituída de X valores em vez de indivíduos ou objetos. A fdp é, então, um modelo para distribuição dos valores da população numérica e, pelo modelo, pode-se calcular diversas características da população (como a média).

Exercícios | Seção 4.1 (1–10)

1. Seja X o tempo que um livro de uma reserva de duas horas, na biblioteca de uma faculdade, é examinado por um estudante selecionado aleatoriamente e suponha que X tenha função de densidade

$$f(x) = \begin{cases} 0{,}5x & 0 \leq x \leq 2 \\ 0 & \text{caso contrário} \end{cases}$$

Calcule as probabilidades a seguir:

a. $P(X \leq 1)$
b. $P(0{,}5 \leq X \leq 1{,}5)$
c. $P(1{,}5 < X)$

2. Suponha que a temperatura de reação X (em °C), em um certo processo químico, tenha função de densidade uniforme para $A = -5$ e $B = 5$.

a. Calcule $P(X < 0)$.
b. Calcule $P(-2{,}5 < X < 2{,}5)$.
c. Calcule $P(-2 \leq X \leq 3)$.

d. Para k satisfazendo $-5 < k < k + 4 < 5$, calcule $P(k < X < k + 4)$.

3. Suponha que o erro envolvido ao se fazer uma certa medida seja uma va contínua X com fdp

$$f(x) = \begin{cases} 0{,}09375(4 - x^2) & -2 \leq x \leq 2 \\ 0 & \text{caso contrário} \end{cases}$$

a. Faça o gráfico de $f(x)$.
b. Calcule $P(X > 0)$.
c. Calcule $P(-1 < X < 1)$.
d. Calcule $P(X < -0{,}5 \text{ ou } X > 0{,}5)$.

4. Seja X a tensão de vibração[1] (psi). Em uma palheta de turbina com certa velocidade de vento em um túnel aerodinâmico. O artigo "Blade Fatigue Life Assessment with Application to VAWTS" (*J. Solar Energy Engr.*, 1982, p. 107–111) propõe a distribuição de Rayleigh, com fdp

$$f(x; \theta) = \begin{cases} \dfrac{x}{\theta^2} \cdot e^{-x^2/(2\theta^2)} & x > 0 \\ 0 & \text{caso contrário} \end{cases}$$

como modelo da distribuição de X.

a. Verifique se $f(x; \theta)$ é uma fdp legítima.
b. Suponha que $\theta = 100$ (valor sugerido: em um gráfico do artigo). Qual é a probabilidade de X ser no máximo 200? Menos que 200? Pelo menos 200?
c. Qual é a probabilidade de X estar entre 100 e 200 (novamente assumindo $\theta = 100$)?
d. Forneça uma expressão para $P(X \leq x)$.

5. Um professor de faculdade nunca finaliza sua aula antes do final do horário e sempre termina dentro de dois minutos após o horário. Seja X = tempo entre o fim do horário e o fim da aula e suponha que a fdp de X seja

$$f(x) = \begin{cases} kx^2 & 0 \leq x \leq 2 \\ 0 & \text{caso contrário} \end{cases}$$

a. Determine o valor de k. [Sugestão: a área total abaixo do gráfico de $f(x)$ é 1].
b. Qual é a probabilidade de a aula terminar dentro de 1 min. do final do horário?
c. Qual é a probabilidade de a aula continuar além do horário por 60 a 90 seg?
d. Qual é a probabilidade de a aula continuar por pelo menos 90 segundos após o final do horário?

6. O peso de tração real de um cartucho de som que está ajustado para rodas com 3g em um certo aparelho pode ser considerado uma va contínua X com fdp

$$f(x) = \begin{cases} k[1 - (x - 3)^2] & 2 \leq x \leq 4 \\ 0 & \text{caso contrário} \end{cases}$$

a. Faça o gráfico de $f(x)$.
b. Determine o valor de k.
c. Qual é a probabilidade de o peso real de tração ser maior do que o nominal?
d. Qual é a probabilidade de o peso real estar dentro de 0,25g de distância em relação ao nominal?
e. Qual é a probabilidade de o peso real diferir do nominal em mais de 0,5g?

7. Acredita-se que o tempo X (min.) para um assistente de laboratório preparar o equipamento para certo experimento tenha distribuição uniforme com $A = 25$ e $B = 35$.

a. Escreva a fdp de X e faça o gráfico.
b. Qual é a probabilidade de o tempo de preparação exceder 33 minutos?
c. Qual é a probabilidade de o tempo de preparação estar dentro de 2 minutos em relação à média? [*Sugestão*: leia μ no gráfico de $f(x)$.]
d. Para qualquer a tal que $25 < a < a + 2 < 35$, qual é a probabilidade de o tempo de preparação estar entre a e $a + 2$ min.?

8. Para ir trabalhar, primeiro preciso tomar um ônibus perto da minha casa e depois mudar de ônibus. Se o tempo de espera (em minutos) em cada parada tem uma distribuição uniforme com $A = 0$ e $B = 5$, pode ser mostrado que meu tempo total de espera Y tem a fdp

$$f(y) = \begin{cases} \dfrac{1}{25}y & 0 \leq y < 5 \\ \dfrac{2}{5} - \dfrac{1}{25}y & 5 \leq y \leq 10 \\ 0 & y < 0 \text{ ou } y > 10 \end{cases}$$

a. Desenhe o gráfico da fdp de Y.
b. Demonstre que $\int_{-\infty}^{\infty} f(y)\, dy = 1$.
c. Qual é a probabilidade de o tempo total de espera ser no máximo 3 min.?
d. Qual é a probabilidade de o tempo total de espera ser no máximo 8 min.?
e. Qual é a probabilidade de o tempo total de espera estar entre 3 e 8 min.?
f. Qual é a probabilidade de o tempo total de espera ser menos de 2 minutos ou mais de 6 minutos?

9. Considere novamente a fdp X = tempo de avanço fornecida no Exemplo 4.4.

a. Qual é a probabilidade de o tempo de avanço ser no máximo 6 segundos?
b. Mais de 6 segundos? Ao menos 6 segundos?
c. Estar entre 5 e 6 segundos?

10. Uma família de fdps que tem sido usada para aproximar a distribuição de renda, tamanho da população da cidade e tamanho das empresas é a família de curvas de Pareto. A família tem dois parâmetros, k e θ, ambos > 0 e a fdp é

$$f(x; k, \theta) = \begin{cases} \dfrac{k \cdot \theta^k}{x^{k+1}} & x \geq \theta \\ 0 & x < \theta \end{cases}$$

a. Faça o gráfico de $f(x; k, \theta)$.
b. Demonstre que a área total abaixo do gráfico é igual a 1.
c. Se a va X tiver fdp $f(x; k, \theta)$, para qualquer $b > \theta$, fixo obtenha uma expressão para $P(X \leq b)$.
d. Para $\theta < a < b$, obtenha uma expressão para a probabilidade $P(a \leq X \leq b)$.

[1] psi é a abreviatura do *pounds per square inch* (libras por polegada quadrada). (NT)

4.2 Funções de Distribuição Acumulada e Valores Esperados

Muitos dos conceitos mais importantes apresentados no estudo de distribuições discretas também desempenham papel importante nas distribuições contínuas. Definições análogas às do Capítulo 3 envolvem a substituição da somatória pelo símbolo de integral.

Função Distribuição Acumulada

A função distribuição acumulada (fdc) $F(x)$ de uma va discreta X fornece, para qualquer número específico x, a probabilidade $P(X \leq x)$. Ela é obtida somando-se fmp $p(y)$ para todos os valores possíveis y que satisfaçam $y \leq x$. A fdc de uma va contínua fornece as mesmas probabilidades $P(X \leq x)$ e é obtida pela integração da fdp $f(y)$ entre os limites $-\infty$ e x.

DEFINIÇÃO

A **função distribuição acumulada** $F(x)$ de uma va contínua X é definida para cada número x por

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy$$

Para cada x, $F(x)$ é a área abaixo da curva de densidade à esquerda de x. Essa propriedade é ilustrada na Figura 4.5, onde $F(x)$ aumenta com ajuste à medida que x aumenta.

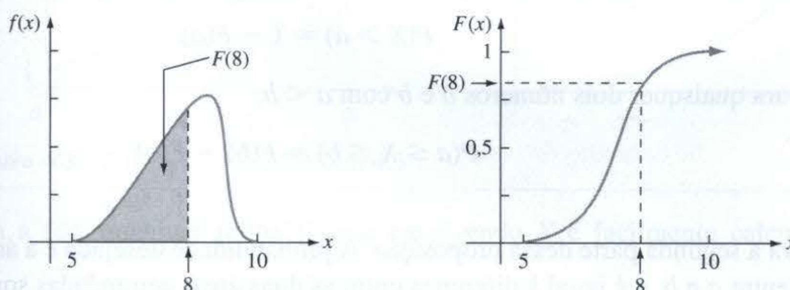


Figura 4.5 Uma fdp e a fdc associada

Exemplo 4.5

Seja X, a espessura de uma determinada chapa de metal, com distribuição uniforme no intervalo $[A, B]$. A função de densidade é exibida na Figura 4.6. Para $x < A$, $F(x) = 0$, porque não há área sob o gráfico da função de densidade à esquerda de x. Para $x \geq B$, $F(x) = 1$, já que toda a área está acumulada à esquerda de x. Finalmente, para $A \leq x \leq B$,

$$F(x) = \int_{-\infty}^{x} f(y)\, dy = \int_{A}^{x} \frac{1}{B-A}\, dy = \frac{1}{B-A} \cdot y \Bigg|_{y=A}^{y=x} = \frac{x-A}{B-A}$$

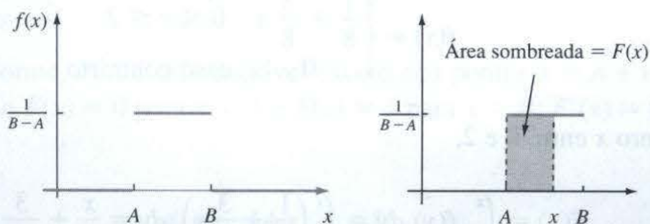


Figura 4.6 A fdp para distribuição uniforme

A fdc completa é

$$F(x) = \begin{cases} 0 & x < A \\ \dfrac{x - A}{B - A} & A \leq x < B \\ 1 & x \geq B \end{cases}$$

O gráfico desta fdc é exibido na Figura 4.7.

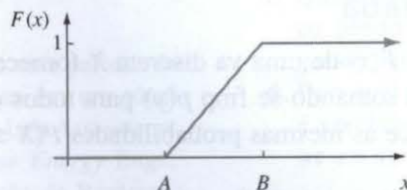


Figura 4.7 A fdc de distribuição uniforme ■

Uso de $F(x)$ para Calcular as Probabilidades

A importância da fdc neste caso, assim como no de vas discretas, é que as probabilidades dos diversos intervalos podem ser calculadas por uma fórmula ou lidas de uma tabela de $F(x)$.

PROPOSIÇÃO Seja X uma va contínua com fdp $f(x)$ e fdc $F(x)$. Então, para qualquer número a,

$$P(X > a) = 1 - F(a)$$

e, para quaisquer dois números a e b com $a < b$,

$$P(a \leq X \leq b) = F(b) - F(a)$$

A Figura 4.8 ilustra a segunda parte dessa proposição. A probabilidade desejada é a área sombreada abaixo da curva de densidade entre a e b, e é igual à diferença entre as duas áreas acumuladas sombreadas.

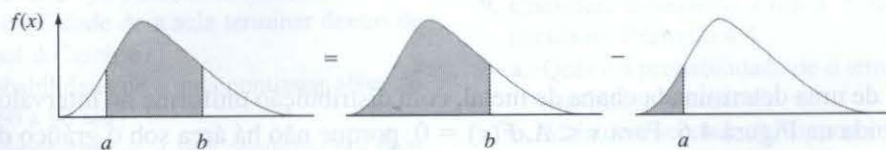


Figura 4.8 Calculo de $P(a \leq X \leq b)$ pelas de probabilidades acumuladas

Exemplo 4.6

Suponha que a fdp da grandeza X de uma carga dinâmica em uma ponte (em newtons) seja dada por

$$f(x) = \begin{cases} \dfrac{1}{8} + \dfrac{3}{8}x & 0 \leq x \leq 2 \\ 0 & \text{caso contrário} \end{cases}$$

Para qualquer número x entre 0 e 2,

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_{0}^{x} \left(\frac{1}{8} + \frac{3}{8}y\right) dy = \frac{x}{8} + \frac{3}{16}x^2$$

Dessa forma,

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{8} + \dfrac{3}{16}x^2 & 0 \le x \le 2 \\ 1 & 2 < x \end{cases}$$

Os gráficos de $f(x)$ e $F(x)$ são exibidos na Figura 4.9. A probabilidade de a carga estar entre 1 e 1,5 é

$$P(1 \le X \le 1{,}5) = F(1{,}5) - F(1)$$
$$= \left[\frac{1}{8}(1{,}5) + \frac{3}{16}(1{,}5)^2\right] - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right]$$
$$= \frac{19}{64} = 0{,}297$$

A probabilidade de a carga exceder 1 é

$$P(X > 1) = 1 - P(X \le 1) = 1 - F(1) = 1 - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right]$$
$$= \frac{11}{16} = 0{,}688$$

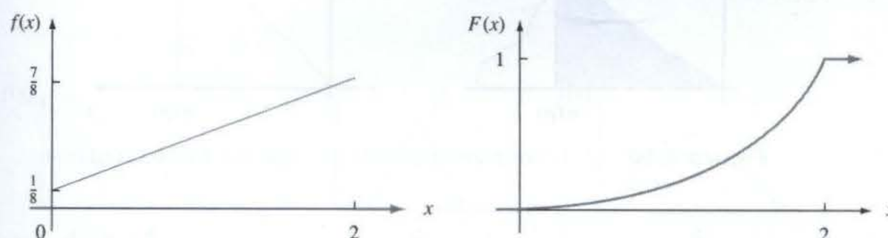


Figura 4.9 A fdp e a fdc do Exemplo 4.6 ■

Uma vez obtida a fdc, qualquer probabilidade envolvendo X é facilmente calculada sem necessidade de integração.

Obtendo $f(x)$ a partir de $F(x)$

Para X discreta, a fmp é obtida pela fdc calculando-se a diferença entre dois valores $F(x)$. A expressão contínua análoga de uma diferença é a derivada. O resultado a seguir é uma conseqüência do Teorema Fundamental do Cálculo.

PROPOSIÇÃO Se X for uma va contínua com fdp $f(x)$ e fdc $F(x)$ então, para qualquer x em que a derivada $F'(x)$ existir, $F'(x) = f(x)$.

Exemplo 4.7 (Continuação do Exemplo 4.5)

Quando X tem distribuição uniforme, $F(x)$ é diferenciável exceto nos pontos $x = A$ e $x = B$, onde o gráfico de $F(x)$ forma cantos agudos. Como $F(x) = 0$ para $x < A$ e $F(x) = 1$ para $x > B$, $F'(x) = 0 = f(x)$ para tal x. Para $A < x < B$,

$$F'(x) = \frac{d}{dx}\left(\frac{x - A}{B - A}\right) = \frac{1}{B - A} = f(x)$$

■

Percentis da Distribuição Contínua

Quando dizemos que a pontuação do teste de um indivíduo estava no 85° percentil da população, queremos dizer que 85% de todas as pontuações da população estavam abaixo daquela pontuação e 15% acima dela. De forma similar, o 40° percentil é a pontuação que excede 40% de todas as pontuações e é excedido por 60%.

DEFINIÇÃO Seja p um número entre 0 e 1. O **100p-ésimo percentil** da distribuição de uma va contínua X, representada por $\eta(p)$ é definido por

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)\, dy \qquad (4.2)$$

De acordo com a Expressão (4.2), $\eta(p)$ é o valor no eixo das medidas tal que $100p\%$ da área abaixo do gráfico de $f(x)$ está à esquerda de $\eta(p)$ e $100(1 - p)\%$ está à direita. Portanto, $\eta(0{,}75)$, o 75° percentil, é tal que a área abaixo do gráfico de $f(x)$ à esquerda de $\eta(0{,}75)$ é 0,75. A Figura 4.10 ilustra a definição.

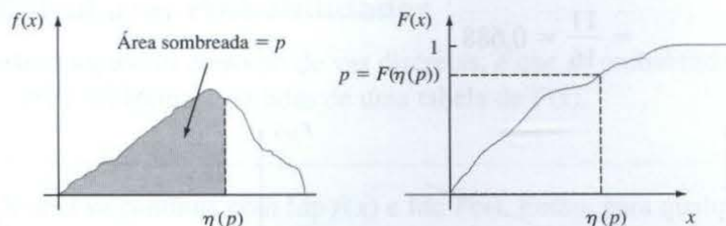


Figura 4.10 O (100p)-ésimo percentil de uma distribuição contínua

Exemplo 4.8

A distribuição da quantidade de cascalho (em toneladas) vendida por uma empresa de materiais de construção em uma semana é uma va contínua X com fdp

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

A fdc de vendas para qualquer x entre 0 e 1 é

$$F(x) = \int_0^x \frac{3}{2}(1 - y^2)\, dy = \frac{3}{2}\left(y - \frac{y^3}{3}\right)\Bigg|_{y=0}^{y=x} = \frac{3}{2}\left(x - \frac{x^3}{3}\right)$$

Os gráficos de $f(x)$ e $F(x)$ são exibidos na Figura 4.11. O (100p)-ésimo percentil dessa distribuição satisfaz a equação

$$p = F(\eta(p)) = \frac{3}{2}\left[\eta(p) - \frac{(\eta(p))^3}{3}\right]$$

isto é,

$$(\eta(p))^3 - 3\eta(p) + 2p = 0$$

Para o 50° percentil, $p = 0{,}5$ e a equação a ser resolvida é $\eta^3 - 3\eta + 1 = 0$; a solução é $\eta = \eta(0{,}5) = 0{,}347$. Se a distribuição continuar igual de semana para semana, no longo prazo, 50% de todas as semanas resultarão em vendas de menos de 0,347 toneladas e 50% em mais de 0,347 toneladas.

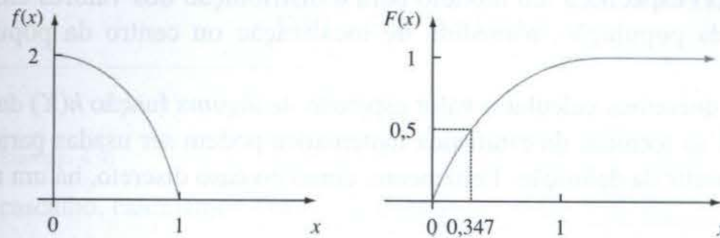


Figura 4.11 A fdp e a fdc do Exemplo 4.8 ■

DEFINIÇÃO

A **mediana** de uma distribuição contínua, representada por $\tilde{\mu}$, é o 50º percentil, de forma que $\tilde{\mu}$ satisfaz $0{,}5 = F(\tilde{\mu})$. Isto é, metade da área abaixo da curva de densidade está à esquerda de $\tilde{\mu}$ e metade à direita.

Uma distribuição contínua cuja fdp é **simétrica** — que significa que o gráfico da fdp à esquerda de um ponto é uma imagem especular do gráfico à direita desse ponto — tem mediana $\tilde{\mu}$ igual ao ponto de simetria, já que metade da área abaixo da curva encontra-se em cada lado desse ponto. A Figura 4.12 fornece diversos exemplos. O valor do erro na medida de uma quantidade física é freqüentemente suposto como tendo distribuição simétrica.

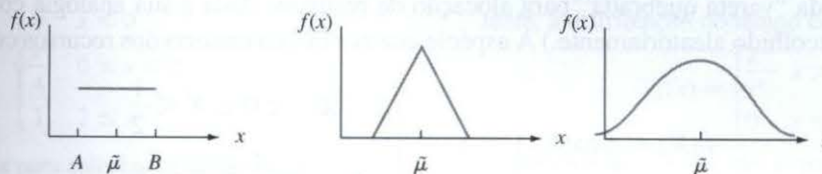


Figura 4.12 Medianas de distribuições simétricas

Valores Esperados de Variáveis Aleatórias Contínuas

Para uma variável aleatória discreta X, $E(X)$ foi obtido pela soma de $x \cdot p(x)$ para os valores possíveis de X. Neste caso, substituímos a soma pela integração e a fmp pela fdp para obtermos uma média contínua ponderada.

DEFINIÇÃO

O **valor médio** ou **esperado** de uma va contínua X com fdp $f(x)$ é

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Exemplo 4.9 (Continuação do Exemplo 4.8)

A fdp das vendas semanais de cascalho X era

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{caso contrário} \end{cases}$$

então

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_0^1 x \cdot \frac{3}{2}(1 - x^2)\, dx$$

$$= \frac{3}{2}\int_0^1 (x - x^3)\, dx = \frac{3}{2}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Bigg|_{x=0}^{x=1} = \frac{3}{8}$$

■

Quando a fdp $f(x)$ especifica um modelo para a distribuição dos valores em uma população numérica, então μ é a média da população, a medida de localização ou centro da população mais freqüentemente usada.

Freqüentemente queremos calcular o valor esperado de alguma função $h(X)$ da va X. Se imaginarmos $h(X)$ como uma nova va Y, as técnicas da estatística matemática podem ser usadas para deduzir a fdp de Y, e $E(Y)$ pode ser calculada a partir da definição. Felizmente, como no caso discreto, há um modo mais fácil de calcularmos $E[h(X)]$.

PROPOSIÇÃO

Se X for uma va contínua com fdp $f(x)$ e $h(X)$ for qualquer função de X, então

$$E[h(X)] = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx$$

Exemplo 4.10

Duas espécies estão competindo em uma região pelo controle de uma quantidade limitada de um determinado recurso. Seja X = a proporção do recurso controlado pela espécie 1 e suponha que X tem fdp

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{caso contrário} \end{cases}$$

que é uma distribuição uniforme no intervalo [0, 1]. (Em seu livro *Ecological Diversity*, E. C. Pielou denomina o caso como o modelo da "vareta quebrada" para alocação de recursos, dada a sua analogia com a quebra de uma vareta em um ponto escolhido aleatoriamente.) A espécie que controla a maioria dos recursos controla a quantidade

$$h(X) = \max(X, 1 - X) = \begin{cases} 1 - X & \text{se } 0 \le X < \frac{1}{2} \\ X & \text{se } \frac{1}{2} \le X \le 1 \end{cases}$$

O valor esperado controlado pela espécie que tem controle majoritário é portanto

$$E[h(X)] = \int_{-\infty}^{\infty} \max(x, 1 - x) \cdot f(x)\, dx = \int_0^1 \max(x, 1 - x) \cdot 1\, dx$$

$$= \int_0^{1/2} (1 - x) \cdot 1\, dx + \int_{1/2}^{1} x \cdot 1\, dx = \frac{3}{4}$$

■

Para $h(X)$, função linear, $E[h(X)] = E(aX + b) = aE(X) + b$.

Variância da Variável Aleatória Contínua

DEFINIÇÃO

A **variância** de uma variável aleatória contínua X com fdp $f(x)$ e média μ é

$$\sigma_X^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)\, dx = E[(X - \mu)^2]$$

O **desvio padrão** (DP) de X é $\sigma_X = \sqrt{V(X)}$.

Como no caso discreto, σ_X^2 é o quadrado do desvio médio ou esperado em relação à média μ e fornece uma medida da dispersão da distribuição ou da população dos valores de x. A forma mais fácil de calcular σ^2 é usar novamente uma fórmula prática.

PROPOSIÇÃO

$$V(X) = E(X^2) - [E(X)]^2$$

Exemplo 4.11 (Continuação do Exemplo 4.9)

Para X = venda semanal de cascalho, calculamos $E(X) = \frac{3}{8}$. Como

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx = \int_0^1 x^2 \cdot \frac{3}{2}(1 - x^2)\,dx$$

$$= \int_0^1 \frac{3}{2}(x^2 - x^4)\,dx = \frac{1}{5}$$

$$V(X) = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320} = 0{,}059 \quad \text{e} \quad \sigma_X = 0{,}244$$

Quando $h(X)$ é uma função linear e $V(X) = \sigma^2$, $V[h(X)] = V(aX + b) = a^2 \cdot \sigma^2$ e $\sigma_{aX+b} = |a| \cdot \sigma$.

Exercícios | Seção 4.2 (11–25)

11. A fdc da duração da retirada X descrita no Exercício 1 é

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x^2}{4} & 0 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

Use tais condições para calcular os itens a seguir

a. $P(X \le 1)$

b. $P(0{,}5 \le X \le 1)$

c. $P(X > 0{,}5)$

d. A mediana da duração da retirada $\tilde{\mu}$ [resolva $0{,}5 = F(\tilde{\mu})$]

e. $F'(x)$ para obter a função de densidade $f(x)$

f. $E(X)$

g. $V(X)$ e σ_X

h. Se o aluno que retira o livro tem uma quantia a pagar $h(X) = X^2$, quando a duração da retirada é X, calcule a taxa esperada $E[h(X)]$.

12. A fdc de X (erro de medida) do Exercício 3 for

$$F(x) = \begin{cases} 0 & x < -2 \\ \dfrac{1}{2} + \dfrac{3}{32}\left(4x - \dfrac{x^3}{3}\right) & -2 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

a. Calcule $P(X < 0)$.

b. Calcule $P(-1 < X < 1)$.

c. Calcule $P(0{,}5 < X)$.

d. Demonstre que $f(x)$ está de acordo com o fornecido no Exercício 3, pela obtenção de $F'(x)$.

e. Demonstre que $\tilde{\mu} = 0$.

13. O Exemplo 4.4 apresentou o conceito de tempo de avanço no fluxo de tráfego e propôs uma distribuição particular para X = o tempo de avanço entre dois carros consecutivos selecionados aleatoriamente (em segundos). Suponha que, em um ambiente de tráfego diferente, a distribuição do tempo de avanço tenha a forma

$$f(x) = \begin{cases} \dfrac{k}{x^4} & x > 1 \\ 0 & x \le 1 \end{cases}$$

a. Determine o valor de k para o qual $f(x)$ é uma fdp legítima.

b. Obtenha a função de distribuição acumulada.

c. Use a fdc de (b) para determinar a probabilidade de o tempo de avanço exceder 2 segundos e também a probabilidade de ele estar entre 2 e 3 segundos.

d. Obtenha o valor médio do tempo de avanço e seu desvio padrão.

e. Qual é a probabilidade de o tempo de avanço estar dentro de 1 desvio padrão em relação à média?

14. O artigo "Modeling Sediment and Water Column Interactions for Hydrophobic Pollutants" (*Water Research*, 1984, p. 1169-1174) sugere a distribuição uniforme no intervalo (7,5, 20) como modelo para a profundidade (cm) da camada de bioturbação em sedimentos de uma determinada região.

a. Qual é a média e a variância da profundidade?

b. Qual é a fdc da profundidade?

c. Qual é a probabilidade de a profundidade observada ser no máximo 10? Entre 10 e 15?

d. Qual é a probabilidade de a profundidade observada estar dentro de 1 desvio padrão em relação ao valor da média? E dentro de 2 desvios padrão?

15. Seja X o espaço ocupado por um produto colocado em um recipiente de 1 pé cúbico. A fdp de X é

$$f(x) = \begin{cases} 90x^8(1 - x) & 0 < x < 1 \\ 0 & \text{caso contrário} \end{cases}$$

a. Desenhe o gráfico da fdp. Determine então a fdc de X e desenhe seu gráfico.
b. Qual é a $P(X \leq 0{,}5)$ [isto é, $F(0{,}5)$]?
c. Usando o item (a), qual é $P(0{,}25 < X \leq 0{,}5)$? Qual é $P(0{,}25 \leq X \leq 0{,}5)$?
d. Qual é o 75º percentil da distribuição?
e. Calcule $E(X)$ e σ_X.
f. Qual é a probabilidade de X estar a mais de 1 desvio padrão em relação ao valor da média?

16. Responda aos itens (a)-(f) do Exercício 15 com X = tempo de palestra além da hora dada no Exercício 5.

17. Considere a fdp de X = peso de tração real fornecido no Exercício 6.
a. Determine e desenhe o gráfico de X.
b. Determine $\tilde{\mu}$ pelo gráfico de $f(x)$.
c. Calcule $E(X)$ e $V(X)$.

18. Considere X tendo distribuição uniforme no intervalo $[A, B]$.
a. Determine uma expressão para o $(100p)$-ésimo percentil.
b. Calcule $E(X)$, $V(X)$, e σ_X.
c. Calcule $E(X^n)$ para n inteiro positivo.

19. Seja X uma va contínua com fdc.

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \dfrac{x}{4}\left[1 + \ln\left(\dfrac{4}{x}\right)\right] & 0 < x \leq 4 \\ 1 & x > 4 \end{cases}$$

[Esse tipo de fdc é sugerido no artigo "Variability in Measured Bedload-Transport Rates" (*Water Resources Bull.*, 1985, p. 39-48) como modelo para determinada variável hidrológica]. Qual é
a. $P(X \leq 1)$?
b. $P(1 \leq X \leq 3)$?
c. A fdp de X?

20. Considere a fdp do tempo total de espera Y para dois ônibus

$$f(y) = \begin{cases} \dfrac{1}{25}y & 0 \leq y < 5 \\ \dfrac{2}{5} - \dfrac{1}{25}y & 5 \leq y \leq 10 \\ 0 & \text{caso contrário} \end{cases}$$

apresentado no Exercício 8.
a. Calcule e desenhe a fdc de Y. [*Sugestão*: considere separadamente $0 \leq y < 5$ e $5 \leq y \leq 10$ ao calcular $F(y)$. Um gráfico da fdp pode ser útil].
b. Determine uma expressão para o $(100p)$-ésimo percentil. (*Sugestão*: considere separadamente $0 < p < 0{,}5$ e $0{,}5 < p < 1$.)
c. Calcule $E(Y)$ e $V(Y)$. Como elas se comparam ao tempo de espera estimado e à variância de um único ônibus quando o tempo tem distribuição uniforme no intervalo $[0, 5]$?

21. Um ecologista deseja marcar uma região de amostragem circular com raio de 10 m. Entretanto, o raio da região resultante é, na verdade, uma variável aleatória R com fdp

$$f(r) = \begin{cases} \dfrac{3}{4}[1 - (10 - r)^2] & 9 \leq r \leq 11 \\ 0 & \text{caso contrário} \end{cases}$$

Qual é a área esperada da região circular resultante?

22. A demanda semanal de gás propano (em milhares de galões) de uma instalação é uma va X com fdp

$$f(x) = \begin{cases} 2\left(1 - \dfrac{1}{x^2}\right) & 1 \leq x \leq 2 \\ 0 & \text{caso contrário} \end{cases}$$

a. Calcule a fdc de X.
b. Determine uma expressão para o $(100p)$-ésimo percentil. Qual é o valor de $\tilde{\mu}$?
c. Calcule $E(X)$ e $V(X)$.
d. Se 1,5 mil galões estão em estoque no começo da semana e nenhuma entrega está programada para a semana, quanto desse volume se espera ter no fim da semana? *Sugestão*: seja $h(x)$ = quantidade restante quando a demanda = x.

23. Se a temperatura de fusão de certo componente for uma variável aleatória com valor médio 120 °C e desvio padrão 2 °C, quais são a temperatura média e o desvio padrão em °F? (*Sugestão*: °F = 1,8 °C + 32.)

24. Seja X com fdp de Pareto

$$f(x; k, \theta) = \begin{cases} \dfrac{k \cdot \theta^k}{x^{k+1}} & x \geq \theta \\ 0 & x < \theta \end{cases}$$

apresentado no Exercício 10.
a. Se $k > 1$, calcule $E(X)$.
b. O que se pode dizer sobre $E(X)$ se $k = 1$?
c. Se $k > 2$, mostre que $V(X) = k\theta^2(k - 1)^{-2}(k - 2)^{-1}$.
d. Se $k = 2$, o que pode ser dito sobre $V(X)$?
e. Que condições em k são necessárias para assegurar que $E(X^n)$ seja finita?

25. Seja X a temperatura em °C em que ocorre uma reação química e seja Y a temperatura em °F (então $Y = 1{,}8X + 32$).
a. Se a mediana da distribuição X é $\tilde{\mu}$, demonstre que $1{,}8\tilde{\mu} + 32$ é a mediana da distribuição Y.
b. Como o 90º percentil da distribuição Y se relaciona com o 90º percentil da distribuição X? Demonstre suas conjecturas.
c. De maneira geral, se $Y = aX + b$, como qualquer percentil da distribuição Y se relaciona com o percentil correspondente da distribuição X?

4.3 A Distribuição Normal

A distribuição normal é a mais importante de todas em probabilidade e estatística. Muitas populações numéricas possuem distribuições que podem ser ajustadas aproximadamente por uma curva normal apropriada. Os exemplos incluem alturas, pesos e outras características físicas (o famoso artigo da *Biometrika*, de 1903 "On the Laws of Inheritance in Man", discutiu muitos exemplos desse tipo), erros de medida em experimentos científicos, medidas antropométricas em fósseis, tempos de reação em experimentos psicológicos, medidas de inteligência e aptidão, pontuações em testes variados e numerosas medidas e indicadores econômicos. Mesmo quando a distribuição em questão é discreta, a curva normal freqüentemente fornece aproximação excelente. Além disso, ainda que as próprias variáveis individuais não sejam normalmente distribuídas, as somas e as médias das variáveis terão uma distribuição aproximadamente normal sob condições adequadas. Essa é a essência do Teorema do Limite Central discutido no próximo Capítulo.

DEFINIÇÃO

Diz-se que uma va contínua X possui uma **distribuição normal** com parâmetros μ e σ (ou μ e σ^2), onde $-\infty < \mu < \infty$ e $0 < \sigma$, se a fdp de X for

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty \tag{4.3}$$

Novamente, e denota a base do sistema de logaritmos naturais e é aproximadamente igual a 2,71828 e π representa a constante matemática familiar com valor aproximado 3,14159. A afirmação de que X possui distribuição normal com parâmetros μ e σ^2 é normalmente abreviada por $X \sim N(\mu, \sigma^2)$.

Claramente $f(x; \mu, \sigma) \geq 0$ para qualquer número x, mas as técnicas do cálculo de multivariados são usadas para demonstrar que $\int_{-\infty}^{\infty} f(x; \mu, \sigma)\, dx = 1$. Demonstra-se que $E(X) = \mu$ e $V(X) = \sigma^2$, de forma que os parâmetros são a média e o desvio padrão de X. A Figura 4.13 apresenta gráficos de $f(x; \mu, \sigma)$ para diversos pares (μ, σ). Cada gráfico é simétrico em torno de μ e tem forma de sino, de modo que o centro do sino (ponto de simetria) é tanto a média como a mediana da distribuição. O valor de σ é a distância de μ até os pontos de inflexão da curva (os pontos em que a curva muda de direção). Valores grandes de σ geram gráficos com grande dispersão em torno de μ, enquanto valores pequenos de σ fornecem gráficos com picos altos acima de μ e a maior parte da área do gráfico é muito próxima de μ. Assim, σ grande implica boa possibilidade de se observar um valor de X afastado de μ, enquanto essa possibilidade é muito improvável para σ pequeno.

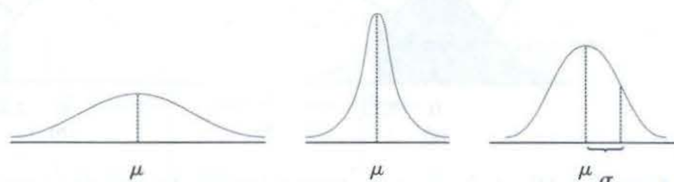


Figura 4.13 Gráficos de fdps normais

Distribuição Normal Padrão

Para calcular $P(a \leq X \leq b)$ quando X é uma va normal com parâmetros μ e σ, devemos calcular

$$\int_a^b \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)}\, dx \tag{4.4}$$

Nenhuma das técnicas de integração-padrão podem ser usadas para calcular a Expressão (4.4). Em vez disso, quando $\mu = 0$ e $\sigma = 1$, a Expressão (4.4) foi calculada numericamente e tabulada para valores determinados de a e b. Essa tabela também é usada para calcular probabilidades de quaisquer outros valores de μ e σ que estejam em consideração.

DEFINIÇÃO A distribuição normal com parâmetros $\mu = 0$ e $\sigma = 1$ é denominada **distribuição normal padrão**. Uma variável aleatória que tenha distribuição normal padrão é denominada **variável aleatória normal padrão** e é denotada por Z. A fdp de Z é

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < \infty$$

A fdc de Z é $P(Z \leq z) = \int_{-\infty}^{z} f(y; 0, 1)$, que será denotada por $\Phi(z)$.

A distribuição normal padrão freqüentemente não serve como modelo para uma população natural. Ao contrário, é uma distribuição de referência a partir da qual podem ser obtidas informações sobre outras distribuições normais. A Tabela A.3 do Apêndice fornece $\Phi(z) = P(Z \leq z)$, a área abaixo do gráfico da fdp normal padrão à esquerda de z, para $z = -3{,}49, -3{,}48, \ldots, 3{,}48, 3{,}49$. A Figura 4.14 ilustra o tipo de área acumulada (probabilidade) tabulada na Tabela A.3. Pela tabela podem ser calculadas diversas outras probabilidades que envolvem Z.

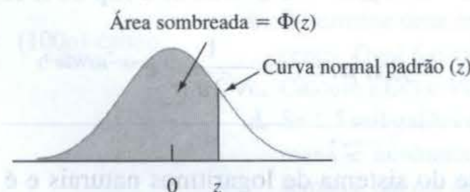


Figura 4.14 Áreas normais padrão acumuladas tabuladas na Tabela A.3 do Apêndice

Exemplo 4.12

Calcule as seguinte probabilidades normais padrão: (a) $P(Z \leq 1{,}25)$, (b) $P(Z > 1{,}25)$, (c) $P(Z \leq -1{,}25)$ e (d) $P(-0{,}38 \leq Z \leq 1{,}25)$.

a. $P(Z \leq 1{,}25) = \Phi(1{,}25)$, probabilidade tabulada na Tabela A.3 do Apêndice na interseção da linha 1,2 com a coluna 0,05. O número mostrado é 0,8944, então $P(Z \leq 1{,}25) = 0{,}8944$. Consulte a Figura 4.15(a).

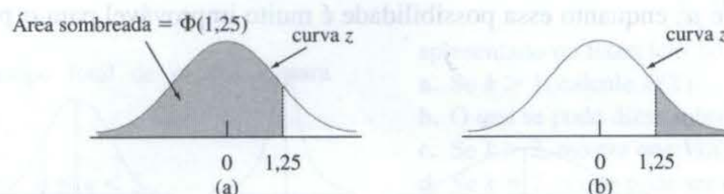


Figura 4.15 Áreas da curva normal (probabilidades) do Exemplo 4.12

b. $P(Z > 1{,}25) = 1 - P(Z \leq 1{,}25) = 1 - \Phi(1{,}25)$, a área abaixo da curva normal padrão à direita de 1,25 (área sob a cauda superior). Como $\Phi(1{,}25) = 0{,}8944$, então $P(Z > 1{,}25) = 0{,}1056$. Como Z é uma va contínua, $P(Z \geq 1{,}25)$ também é igual a 0,1056. Consulte a Figura 4.15 (b).

c. $P(Z \leq -1{,}25) = \Phi(-1{,}25)$, a área sob a cauda inferior. Diretamente da Tabela A.3 do Apêndice, $\Phi(-1{,}25) = 0{,}1056$. Por simetria da curva normal, esta é a mesma resposta do item (b).

d. $P(-0{,}38 \leq Z \leq 1{,}25)$ é a área abaixo da curva normal padrão acima do intervalo cujo ponto extremo esquerdo é $-0{,}38$ e o direito é 1,25. Da Seção 4.2, se X é uma va contínua com fdc $F(x)$, então $P(a \leq X \leq b) = F(b) - F(a)$. O que fornece $P(-0{,}38 \leq Z \leq 1{,}25) = \Phi(1{,}25) - \Phi(-0{,}38) = 0{,}8944 - 0{,}3520 = 0{,}5424$. (Consulte a Figura 4.16.)

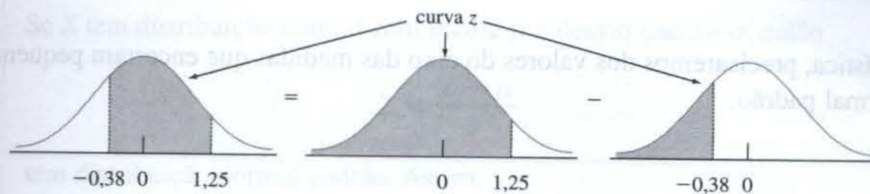


Figura 4.16 $P(-0,38 \leq Z \leq 1,25)$ como a diferença entre duas áreas acumuladas ■

Percentis da Distribuição Normal Padrão

Para qualquer p entre 0 e 1, a Tabela A.3 do Apêndice é usada para obter o $(100p)$-ésimo percentil da distribuição normal padrão.

Exemplo 4.13

O 99º percentil da distribuição normal padrão é o valor no eixo horizontal tal que a área sob a curva à esquerda do valor seja 0,9900. A Tabela A.3 do Apêndice fornece para z fixo a área abaixo da curva normal à esquerda de z, neste caso temos a área e queremos o valor de z. Este é o problema "inverso" de $P(Z \leq z) = ?$ de modo que a tabela é usada de maneira inversa: encontre na tabela o valor 0,9900; a linha e a coluna onde se acha o valor identificam o 99º percentil de z. No caso, 0,9901 está na linha 2,3 e coluna 0,03, então o 99º percentil é (aproximadamente) $z = 2,33$. (Veja a Figura 4.17.) Por simetria, o primeiro percentil é o negativo do 99º percentil, então é igual a $-2,33$ (1% fica abaixo do primeiro e acima do 99º). (Veja a Figura 4.18.)

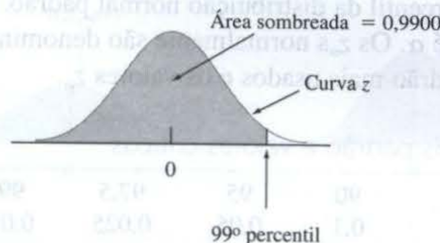


Figura 4.17 Determinação do 99º percentil

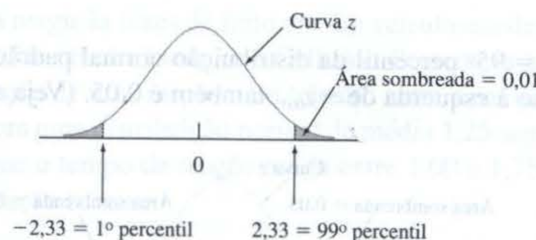


Figura 4.18 Relação entre o 1º e o 99º percentis ■

Em geral, o $(100p)$ésimo percentil é identificado pela linha e coluna da Tabela A.3 do Apêndice onde se encontra a entrada p (por exemplo: o 67º percentil é obtido encontrando-se 0,6700 no corpo da tabela, que fornece $z = 0,44$). Se p não estiver na tabela, normalmente é usado o número mais próximo a ele, apesar de que a interpolação linear fornece uma resposta mais precisa. Por exemplo: para determinar o 95º percentil, procuramos 0,9500 na tabela. Apesar de 0,9500 não aparecer, os valores 0,9495 e 0,9505 são mostrados e correspondem a $z = 1,64$ e 1,65, respectivamente. Como 0,9500 está na metade da diferença entre as duas probabilidades mostradas, usaremos 1,645 como o 95º percentil e $-1,645$ como o 5º percentil.

Notação z_α

Em inferência estatística, precisaremos dos valores do eixo das medidas que encerram pequenas áreas da cauda abaixo da curva normal padrão.

> **Notação z_α**
>
> z_α representará o valor no eixo das medidas para o qual uma área abaixo da curva z fica à direita de z_α. (Veja a Figura 4.19.)

Por exemplo: $z_{0,10}$ contém a área sob a cauda superior 0,10 e $z_{0,01}$ contém a área sob a cauda superior 0,01.

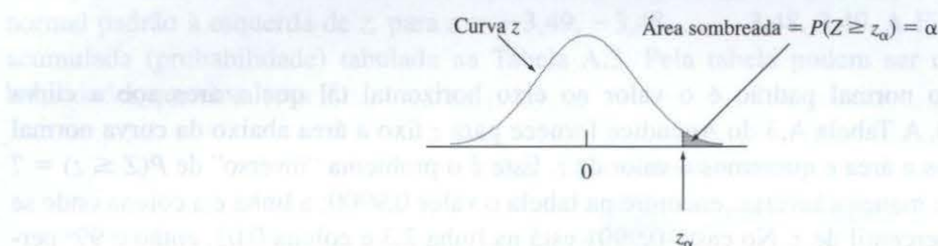


Figura 4.19 Notação z_α ilustrada

Como α da área abaixo da curva normal padrão está à direita de z_α, $1 - \alpha$ da área está à esquerda de z_α. Portanto, z_α é o $100(1 - \alpha)$-ésimo percentil da distribuição normal padrão. Por simetria, a área abaixo da curva normal e à esquerda de $-z_\alpha$ também é α. Os z_αs normalmente são denominados **valores críticos de** z. A Tabela 4.1 relaciona os percentis normais padrão mais usados e os valores z_α.

Tabela 4.1 Percentis normais padrão e valores críticos

Percentil	90	95	97,5	99	99,5	99,9	99,95
α (área sob a cauda)	0,1	0,05	0,025	0,01	0,005	0,001	0,0005
$z_\alpha = 100(1 - \alpha)$-ésimo percentil	1,28	1,645	1,96	2,33	2,58	3,08	3,27

Exemplo 4.14

$z_{0,05}$ é o $100(1 - 0,05)$ésimo = 95º percentil da distribuição normal padrão, de modo que $z_{0,05} = 1,645$. A área abaixo da curva normal padrão à esquerda de $-z_{0,05}$ também é 0,05. (Veja a Figura 4.20). ■

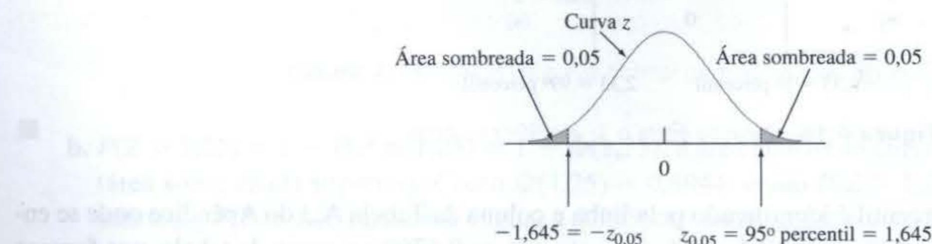


Figura 4.20 Determinação de $z_{0,05}$

Distribuições Normais Não-Padrão

Quando $X \sim N(\mu, \sigma^2)$, as probabilidades que envolvem X são calculadas por "padronização". A **variável padrão** é $(X - \mu)/\sigma$. A subtração de μ desloca a média de μ para zero e depois a divisão por σ escalona a variável, de forma que o desvio padrão passa a ser 1 em vez de σ.

PROPOSIÇÃO

Se X tem distribuição normal com média μ e desvio padrão σ, então

$$Z = \frac{X - \mu}{\sigma}$$

tem distribuição normal padrão. Assim,

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(X \leq a) = \Phi\left(\frac{a - \mu}{\sigma}\right) \quad P(X \geq b) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right)$$

A idéia-chave da proposição é que, pela padronização, qualquer probabilidade que envolva X é expressa como uma probabilidade que envolve uma va normal padrão Z, de forma que a Tabela A.3 do Apêndice pode ser usada. A figura 4.21 ilustra o caso. A proposição é demonstrada escrevendo a fdc de $Z = (X - \mu)/\sigma$ como

$$P(Z \leq z) = P(X \leq \sigma z + \mu) = \int_{-\infty}^{\sigma z + \mu} f(x; \mu, \sigma)\, dx$$

Usando um resultado do cálculo, integral pode ser diferenciada em relação à z para fornecer a fdp desejada $f(z; 0, 1)$.

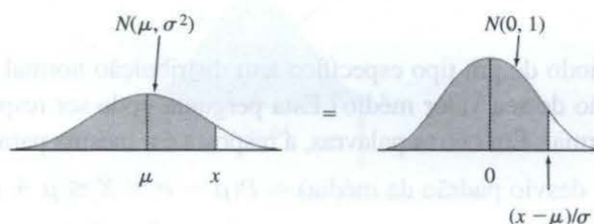


Figura 4.21 Igualdade das áreas de curvas normais não-padrão e padrão

Exemplo 4.15

O tempo que um motorista leva para reagir às luzes de freio em um veículo em desaceleração é crucial para evitar colisões traseiras. O artigo "Fast-Rise Brake Lamp as a Collision-Prevention Device" (*Ergonomics*, 1993, p. 391-395) sugere que o tempo de reação de uma resposta no trânsito a um sinal de frenagem com luzes de freio convencionais pode ser modelado com uma distribuição normal de média 1,25 segundo e desvio padrão 0,46 segundo. Qual é a probabilidade de que o tempo de reação esteja entre 1,00 e 1,75 segundo? Se representarmos por X o tempo de reação, a padronização fornece

$$1{,}00 \leq X \leq 1{,}75$$

se, e somente se,

$$\frac{1{,}00 - 1{,}25}{0{,}46} \leq \frac{X - 1{,}25}{0{,}46} \leq \frac{1{,}75 - 1{,}25}{0{,}46}$$

Dessa forma,

$$p(1{,}00 \leq X \leq 1{,}75) = P\left(\frac{1{,}00}{0{,}46} \leq Z \leq \frac{1{,}75 - 1{,}25}{0{,}46}\right)$$

$$= P(-0{,}54 \leq Z \leq 1{,}09) = \Phi(1{,}09) - \Phi(-0{,}54)$$

$$= 0{,}8621 - 0{,}2946 = 0{,}5675$$

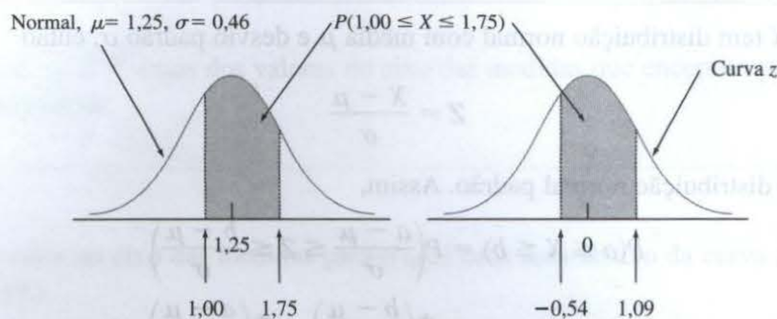


Figura 4.22 Curvas normais do Exemplo 4.15

A Figura 4.22 ilustra o problema. De forma similar, se considerarmos que 2 segundos é um tempo de resposta muito longo, a probabilidade de que o tempo real de resposta exceda esse valor será

$$P(X > 2) = P\left(Z > \frac{2 - 1{,}25}{0{,}46}\right) = P(Z > 1{,}63) = 1 - \Phi(1{,}63) = 0{,}0516$$

■

A padronização nada mais é que calcular uma distância em relação à média e então expressar essa distância novamente como certo número de desvios padrão. Por exemplo: se $\mu = 100$ e $\sigma = 15$, então $x = 130$ corresponde a $z = (130 - 100)/15 = 30/15 = 2{,}00$. Isto é, 130 está a 2 desvios padrão acima (à direita) do valor da média. De forma similar, padronizar 85 fornece $(85 - 100)/15 = -1{,}00$, de forma que 85 está 1 desvio padrão abaixo da média. A tabela z se aplica a qualquer distribuição normal, desde que pensemos em termos de número de desvios padrão em relação à média.

Exemplo 4.16

A voltagem de quebra de um diodo de um tipo específico tem distribuição normal. Qual é a probabilidade dessa voltagem estar a 1 desvio padrão de seu valor médio? Esta pergunta pode ser respondida sem conhecer μ ou σ, desde que a distribuição seja normal. Em outras palavras, a resposta é a mesma para qualquer distribuição normal:

$$\begin{aligned} P(X \text{ está no intervalo de 1 desvio padrão da média}) &= P(\mu - \sigma \leq X \leq \mu + \sigma) \\ &= P\left(\frac{\mu - \sigma - \mu}{\sigma} \leq Z \leq \frac{\mu + \sigma - \mu}{\sigma}\right) \\ &= P(-1{,}00 \leq Z \leq 1{,}00) \\ &= \Phi(1{,}00) - \Phi(-1{,}00) = 0{,}6826 \end{aligned}$$

A probabilidade de X estar no intervalo de 2 desvios-padrão é $P(-2{,}00 \leq Z \leq 2{,}00) = 0{,}9544$ e no intervalo de 3 desvios padrão é $P(-3{,}00 \leq Z \leq 3{,}00) = 0{,}9974$. ■

Os resultados do Exemplo 4.16 são freqüentemente relatados sob a forma de porcentagem e são denominados *regra empírica* (porque evidências empíricas mostram que histogramas de dados reais freqüentemente podem ser aproximados por curvas normais).

> Se a distribuição de população de uma variável for (aproximadamente) normal, então
>
> 1. Cerca de 68% dos valores estão a 1 DP da média.
> 2. Cerca de 95% dos valores estão a 2 DPs da média.
> 3. Cerca de 99,7% dos valores estão a 3 DPs da média.

É incomum observar um valor de uma população normal muito mais distante de 2 desvios padrão de μ. Esses resultados serão importantes no desenvolvimento dos procedimentos de teste de hipóteses nos capítulos posteriores.

Percentis de uma Distribuição Normal Arbitrária

O (100p)-ésimo percentil de uma distribuição normal com média μ e desvio padrão σ pode ser facilmente relacionado com o (100p)-ésimo percentil da distribuição normal padrão.

PROPOSIÇÃO

$$\begin{matrix}(100p)\text{-ésimo percentil} \\ \text{para a normal } (\mu, \sigma)\end{matrix} = \mu + \left[\begin{matrix}(100p)\text{-ésimo para a} \\ \text{normal padronizada}\end{matrix}\right] \cdot \sigma$$

Outra forma de expressar a proposição: se z é o percentil desejado da distribuição normal padrão, então o percentil desejado da distribuição normal (μ, σ) está a z desvios padrão de μ.

Exemplo 4.17

A quantidade de água destilada produzida por certa máquina tem distribuição normal com valor médio de 64 onças e desvio padrão de 0,78 onça. Qual o tamanho de recipiente c que assegurará que ocorra transbordamento em apenas 0,5% das vezes? Se X representa a quantidade usada, a condição desejada é que $P(X > c) = 0{,}005$, ou, de forma equivalente, que $P(X \leq c) = 0{,}995$. Assim, c é o 99,5º percentil da distribuição normal com $\mu = 64$ e $\sigma = 0{,}78$. O 99,5º percentil da distribuição normal padrão é 2,58, então

$$c = \eta(0{,}995) = 64 + (2{,}58)(0{,}78) = 64 + 2{,}0 = 66 \text{ onças}$$

A questão está ilustrada na Figura 4.23.

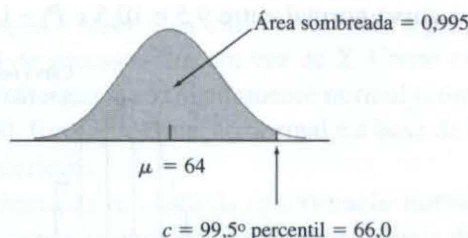


Figura 4.23 Distribuição da quantidade produzida do Exemplo 4.17 ■

A Distribuição Normal e as Populações Discretas

A distribuição normal é usada com freqüência como aproximação para a distribuição de valores de uma população discreta. Em tais situações, devemos tomar cuidado redobrado para nos assegurarmos de que as probabilidades sejam calculadas de forma precisa.

Exemplo 4.18

O QI em uma determinada população (medido pelo teste-padrão) é sabido ter distribuição aproximadamente normal com $\mu = 100$ e $\sigma = 15$. Qual é a probabilidade de um indivíduo selecionado aleatoriamente ter um QI de pelo menos 125? Sendo X = QI de uma pessoa selecionada aleatoriamente, queremos saber $P(X \geq 125)$. A tentação aqui é padronizar $X \geq 125$ imediatamente, como foi feito nos exemplos anteriores. O problema, entretanto, é que a população de QI é discreta, visto que os QI's são valores inteiros, de forma que a curva normal é a aproximação de um histograma de probabilidade discreta, conforme ilustrado na Figura 4.24.

Os retângulos do histograma estão centrados em inteiros, de forma que o QI de pelo menos 125 corresponde aos retângulos começando em 124,5, conforme sombreado na Figura 4.24. Assim, queremos a área abaixo da curva normal de aproximação à direita de 124,5. A padronização desse esse valor fornece $P(Z \geq 1{,}63) = 0{,}0516$. Se tivéssemos padronizado $X \geq 125$, teríamos obtido $P(Z \geq 1{,}67) = 0{,}0475$. A diferença não é grande, mas a resposta 0,0516 é mais precisa. De forma similar, $P(X = 125)$ seria aproximada pela área entre 124,5 e 125,5, já que a área abaixo da curva normal acima do valor 125 é zero.

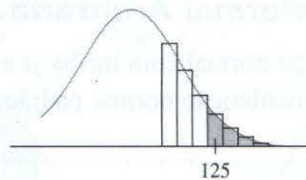


Figura 4.24 Aproximação normal de uma distribuição discreta ■

A correção da descontinuidade da distribuição subjacente no Exemplo 4.18 é freqüentemente denominada **correção de continuidade**. Ela é útil na seguinte aplicação da distribuição normal, para cálculo de probabilidades da binomial.

Aproximação Normal para a Distribuição Binomial

Lembre-se de que o valor médio e o desvio padrão de uma variável aleatória X são $\mu_X = np$ e $\sigma_X = \sqrt{npq}$, respectivamente. A Figura 4.25 exibe um histograma de probabilidade binomial para a distribuição binomial com $n = 20$, $p = 0{,}6$ [de forma que $\mu = 20(0{,}6) = 12$ e $\sigma = \sqrt{20(0{,}6)(0{,}04)} = 2{,}19$]. Uma curva normal com valor médio e desvio padrão iguais aos valores correspondentes para a distribuição binomial foi sobreposta no histograma de probabilidade. Apesar de o histograma ter um pouco de inclinação (porque $p \neq 0{,}5$), a curva normal fornece uma aproximação muito boa, especialmente na parte central da figura. A área de qualquer retângulo (probabilidade de qualquer valor X particular), exceto aqueles nas extremidades das caudas, pode ser aproximada com precisão pela área correspondente da curva normal. Por exemplo: $P(X = 10) = B(10; 20, 0{,}6) - B(9; 20, 0{,}6) = 0{,}117$, enquanto a área sob a curva normal entre 9,5 e 10,5 é $P(-1{,}14 \leq Z \leq -0{,}68) = 0{,}1212$.

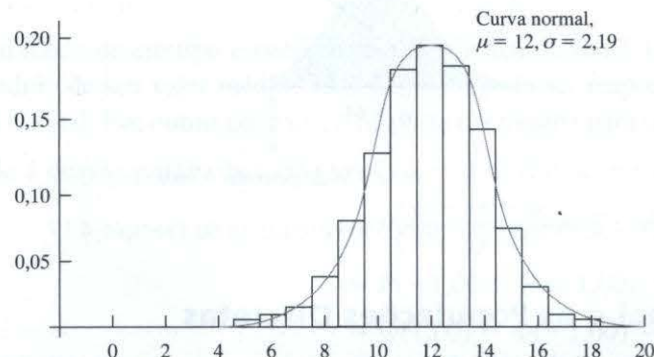


Figure 4.25 Histograma de probabilidade binomial para $n = 20$, $p = 0{,}6$ com curva de aproximação normal sobreposta

De forma mais geral, contanto que o histograma de probabilidade não apresente muita inclinação, as probabilidades binomiais podem ser bem aproximadas pelas áreas da curva normal. Habitualmente se diz que X tem uma distribuição aproximadamente normal.

PROPOSIÇÃO Seja X uma va binomial com base em n tentativas com probabilidade de sucesso p. Então, se o histograma de probabilidade binomial não tiver muita inclinação, X terá uma distribuição aproximadamente normal com $\mu = np$ e $\sigma = \sqrt{npq}$. Em particular, para x = valor possível de X,

$$P(X \leq x) = B(x; n, p) \approx \begin{pmatrix} \text{área abaixo da curva normal} \\ \text{à esquerda de } x + 0{,}5 \end{pmatrix}$$

$$= \Phi\left(\frac{x + 0{,}5 - np}{\sqrt{npq}}\right)$$

Na prática, a aproximação é adequada desde que $np \geq 10$ e $nq \geq 10$.

Se $np < 10$ ou $nq < 10$, a distribuição terá muita inclinação para que a curva normal (simétrica) forneça uma aproximação precisa.

Exemplo 4.19

Suponha que 25% de todos os motoristas habilitados de um determinado estado não tenham seguro. Represente por X o número de motoristas sem seguro em uma amostra aleatória de tamanho 50 (com um quê de perversidade, o sucesso é o motorista não ter seguro), de forma que $p = 0{,}25$. Então $\mu = 12{,}5$ e $\sigma = 3{,}06$. Como $np = 50(0{,}25) = 12{,}5 \geq 10$ e $nq = 37{,}5 \geq 10$, a aproximação pode ser aplicada com segurança:

$$P(X \leq 10) = B(10; 50, 0{,}25) \approx \Phi\left(\frac{10 + 0{,}5 - 12{,}5}{3{,}06}\right)$$

$$= \Phi(-0{,}65) = 0{,}2578$$

De forma similar, a probabilidade de inclusive entre 5 e 15 motoristas selecionados não terem seguro é

$$P(5 \leq X \leq 15) = B(15; 50, 0{,}25) - B(4; 50, 0{,}25)$$

$$\approx \Phi\left(\frac{15{,}5 - 12{,}5}{3{,}06}\right) - \Phi\left(\frac{4{,}5 - 12{,}5}{3{,}06}\right) = 0{,}8320$$

As probabilidades exatas são 0,2622 e 0,8348, respectivamente, de forma que as aproximações são muito boas. No último cálculo, a probabilidade $P(5 \leq X \leq 15)$ é aproximada pela área sob a curva normal entre 4,5 e 15,5. A correção de continuidade é usada para os limites superior e inferior. ■

Quando o objetivo da investigação é fazer uma inferência sobre a proporção p de uma população, o interesse enfocará a proporção amostral de sucessos X/n em vez de X. Como essa proporção é X multiplicada pela constante $1/n$, também terá uma distribuição aproximadamente normal (com média $\mu = p$ e desvio padrão $\sigma = \sqrt{pq/n}$) desde que $np \geq 10$ e $nq \geq 10$. Essa aproximação normal é a base de diversos procedimentos inferenciais a serem discutidos nos capítulos posteriores.

É muito difícil dar uma prova direta da validade da aproximação normal (a primeira é de Laplace, há mais de 150 anos). No próximo capítulo, veremos que ela é uma conseqüência de um importante resultado geral denominado Teorema do Limite Central.

Exercícios | Seção 4.3 (26–54)

26. Seja Z uma variável aleatória normal e calcule as probabilidades a seguir, fazendo as ilustrações quando apropriado.

a. $P(0 \leq Z \leq 2{,}17)$
b. $P(0 \leq Z \leq 1)$
c. $P(-2{,}50 \leq Z \leq 0)$
d. $P(-2{,}50 \leq Z \leq 2{,}50)$
e. $P(Z \leq 1{,}37)$
f. $P(-1{,}75 \leq Z)$
g. $P(-1{,}50 \leq Z \leq 2{,}00)$
h. $P(1{,}37 \leq Z \leq 2{,}50)$
i. $P(1{,}50 \leq Z)$
j. $P(|Z| \leq 2{,}50)$

27. Em cada caso, determine o valor da constante c que torna correta a declaração de probabilidade.

a. $\Phi(c) = 0{,}9838$
b. $P(0 \leq Z \leq c) = 0{,}291$
c. $P(c \leq Z) = 0{,}121$
d. $P(-c \leq Z \leq c) = 0{,}668$
e. $P(c \leq |Z|) = 0{,}016$

28. Encontre os percentis a seguir para a distribuição normal. Faça a interpolação quando apropriado.

a. 91º
b. 9º
c. 75º
d. 25º
e. 6º

29. Determine z_α para os itens a seguir:

a. $\alpha = 0{,}0055$
b. $\alpha = 0{,}09$
c. $\alpha = 0{,}663$

30. Se X é uma va normal com média 80 e desvio padrão 10, calcule as probabilidades a seguir, usando padronização:

a. $P(X \leq 100)$
b. $P(X \leq 80)$
c. $P(65 \leq X \leq 100)$
d. $P(70 \leq X)$
e. $P(85 \leq X \leq 95)$
f. $P(|X - 80| \leq 10)$

31. Suponha que a força que age sobre uma coluna que ajuda a suportar um edifício tenha distribuição normal com média 15,0 kips e desvio padrão 1,25 kips. Qual é a probabilidade de a força:
a. Ser no máximo 18 kips?
b. Estar entre 10 e 12 kips?
c. Diferir de 15,0 kips por no máximo 2 desvios padrão?

32. O artigo "Reliability of Domestic-Waste Biofilm Reactors" (*J. of Emir. Engr.*, 1995, p. 785-790) sugere que a concentração de substrato (mg/cm3) de influência para um reator tem distribuição normal com $\mu = 0{,}30$ e $\sigma = 0{,}06$.
a. Qual é a probabilidade de a concentração exceder 0,25?
b. Qual é a probabilidade de a concentração ser no máximo 0,10?
c. Como você caracterizaria os maiores 5% de todos os valores de concentração?

33. Suponha que o diâmetro de certo tipo de árvores na altura do tronco tenha distribuição normal com $\mu = 8{,}8$ e $\sigma = 2{,}8$, conforme sugerido pelo artigo "Simulating a Harvester-Forwarder Softwood Thinning" (*Forest Products J., May* 1997, p. 36-41).
a. Qual é a probabilidade de uma árvore selecionada aleatoriamente ter um diâmetro de no mínimo 10 polegadas? Exceder 10 polegadas?
b. Qual é a probabilidade de o diâmetro de uma árvore selecionada aleatoriamente exceder 20 polegadas?
c. Qual é a probabilidade de o diâmetro de uma árvore selecionada aleatoriamente estar entre 5 e 10 polegadas?
d. Que valor c faz com que o intervalo $(8{,}8 - c, 8{,}8 + c)$ inclua 98% de todos os valores de diâmetro?
e. Se quatro árvores forem selecionadas de forma independente, qual é a probabilidade de ao menos uma ter diâmetro maior que 10 polegadas?

34. Há duas máquinas disponíveis para corte de rolhas para garrafas de vinho. A primeira produz rolhas com diâmetros que possuem uma distribuição normal com média 3 cm e desvio padrão 0,1 cm. A segunda máquina produz rolhas com diâmetros que possuem uma distribuição normal com média 3,04 cm e desvio padrão 0,02 cm. As rolhas aceitáveis possuem diâmetros entre 2,9 cm e 3,1 cm. Que máquina tem maior probabilidade de produzir uma rolha aceitável?

35. a. Se uma distribuição normal tem $\mu = 30$ e $\sigma = 5$, qual é o 91º percentil da distribuição?
b. Qual é o 6º percentil da distribuição?
c. A largura de uma linha gravada em um chipe de circuito integrado tem distribuição normal com média 3,000 μm e desvio-padrão 0,140. Que valor separa os 10% mais largos de todas as linhas dos outros 90%?

36. O artigo "Monte Carlo Simulation-Tool for Better Understanding of LRFD" (*J. Structural Engr*, 1993, p. 1586-1599) sugere que a resistência de rendimento (ksi) de aço de graduação A36 tem distribuição normal com $\mu = 43$ e $\sigma = 4{,}5$.
a. Qual é a probabilidade de a resistência ser no máximo 40? Maior que 60?
b. Que valor de resistência separa os 75% mais fortes dos outros?

37. O dispositivo de abertura automática de um pára-quedas de carga militar foi projetado para abrir quando estiver a 200 m do solo. Suponha que a altitude de abertura tenha uma distribuição normal com média 200 m e desvio padrão 30 m. Haverá dano no equipamento se o pára-quedas abrir a uma altitude inferior a 100 m. Qual é a probabilidade de haver dano ao equipamento em pelo menos um de cinco pára-quedas lançados independentemente?

38. A leitura de temperatura de um par termoelétrico colocado em um meio de temperatura constante tem distribuição normal com média μ e desvio-padrão σ. Que valor σ precisaria existir para assegurar que 95% de todas as leituras estão dentro de 0,1° de μ?

39. A distribuição da resistência de resistores de um tipo específico é normal, 10% de todos os equipamentos apresentam resistência maior que 10,256 ohms e 5% com resistência menor que 9,671 ohms. Quais são os valores da média e do desvio padrão da distribuição das resistências?

40. Se o comprimento da rosca de um parafuso tem distribuição normal, qual é a probabilidade de o comprimento da rosca de um parafuso selecionado aleatoriamente:
a. estar a 1,5 DPs do valor da média?
b. estar a mais de 2,5 DPs do valor da média?
c. estar entre 1 e 2 DPs do valor da média?

41. Uma máquina que produz rolamentos inicialmente foi configurada para que o diâmetro real médio dos rolamentos produzidos seja de 0,500 polegadas. Um rolamento é aceitável se o diâmetro está dentro de 0,004 polegadas desse valor-alvo. Suponha, entretanto, que a configuração tenha sido alterada durante o curso da produção, de forma que os rolamentos tenham diâmetros com distribuição normal com média 0,499pol e desvio padrão 0,002 polegadas. Que porcentagem dos rolamentos produzidos não será aceitável?

42. A dureza Rockwell de um metal é determinada pela pressão de uma ponta rígida na superfície do metal e, em seguida, pela medição da profundidade de penetração das pontas. Suponha que a dureza Rockwell de uma determinada liga tenha distribuição normal com média 70 e desvio padrão 3. (A dureza Rockwell é medida em uma escala contínua.)

$P(2{,}9 \leq X_1 \leq 3{,}1) = P\left(\frac{2{,}9-\mu_1}{\sigma_1} \leq \frac{X_1-\mu_1}{\sigma_1} \leq \frac{3{,}1-\mu_1}{\sigma_1}\right) = P(-1 \leq Z \leq 1) = F(1) - F(-1) = 96$

a. Se uma amostra for aceitável apenas se a sua dureza estiver entre 67 e 75, qual é a probabilidade de uma amostra selecionadoa aleatoriamente ter dureza aceitável?
b. Se o intervalo aceitável de dureza for $(70 - c, 70 + c)$, para que valor de c 95% de todos das amostras teriam uma dureza aceitável?
c. Se o intervalo aceitável for o da parte (a) e a dureza de cada uma de 10 amostras selecionadas aleatoriamente for determinada de forma independente, qual será o número esperado de amostras aceitáveis dentre os 10?
d. Qual é a probabilidade de no máximo oito de 10 amostras selecionadas de forma independente terem dureza inferior a 73,84? (*Sugestão*: Y = número entre as 10 amostras com dureza inferior a 73,84 é uma variável binomial. Qual é p?)

43. A distribuição do peso de pacotes enviados de uma determinada forma é normal com média 12 libras e desvio padrão 3,5 libras. O serviço de encomendas deseja determinar um valor c acima do qual haverá um acréscimo. Que valor de c faz com que 99% de todos os pacotes estejam no mínimo 1 libra abaixo do peso de acréscimo?

44. Suponha que a Tabela A.3 do Apêndice contenha $\Phi(z)$ apenas para $z \geq 0$. Explique como você pode calcular:
a. $P(-1{,}72 \leq Z \leq -0{,}55)$
b. $P(-1{,}72 \leq Z \leq 0{,}55)$

É necessário tabelar $\Phi(z)$ para z negativo? Que propriedade da curva normal padrão justifica sua resposta?

45. Considere os bebês nascidos no intervalo "normal" de 37-43 semanas de gestação. Muitos dados apóiam a suposição de que, para os bebês nascidos nos Estados Unidos, o peso de nascimento possui distribuição normal com média 3432 g e desvio padrão 482 g. [O artigo "Are Babies Normal" (*The American Statistician*, 1999, p. 298-302) analisou dados de um ano específico. Para uma escolha significativa de intervalos de classe, um histograma não tinha aparência normal, mas, após mais investigações, determinou-se que isso acontecia devido a alguns hospitais medirem o peso em gramas e outros em onças (com aproximação para a próxima onça) e depois converterem para gramas. Uma escolha de intervalos de classe modificada resultou um histograma bem-descrito por uma distribuição normal.]
a. Qual é a probabilidade de o peso de nascimento de um bebê selecionado aleatoriamente exceder 4000 gramas? Estar entre 3000 e 4000 gramas?
b. Qual é a probabilidade de o peso de um bebê selecionado aleatoriamente ser inferior a 2000 gramas ou superior a 5000 gramas?
c. Qual é a probabilidade de o peso de nascimento de um bebê selecionado aleatoriamente exceder 7 libras?
d. Como você caracterizaria o 0,1% mais extremo de todos os pesos de nascimento?
e. Se X for uma variável aleatória com distribuição normal e a for uma constante numérica ($a \neq 0$), então $Y = aX$ também terá uma distribuição normal. Use isso para determinar a distribuição do peso de nascimento expresso em libras (formato, média e desvio padrão) e então calcule novamente a probabilidade da parte (c). Como isso se compara à resposta anterior?

46. Em resposta às preocupações sobre o conteúdo nutricional dos *fast foods*, o McDonald's anunciou que usará um novo óleo na fritura de suas batatas. O produto diminuirá substancialmente os níveis de ácido de gordura trans e aumentará a quantidade de gordura poliinsaturada mais benéfica. A empresa alega que 97 dentre 100 pessoas não conseguem detectar diferença no gosto entre o óleo antigo e o novo. Assumindo que esse valor esteja correto (como proporção de longo prazo), qual é a probabilidade aproximada de, em uma amostra de 1000 indivíduos que compraram fritas no McDonald's,
a. ao menos 40 sentirem a diferença de gosto entre os dois óleos?
b. no máximo 5% sentirem a diferença de gosto entre os dois óleos?

47. A desigualdade de Chebyshev, apresentada no Exercício 43 (Capítulo 3), é válida para distribuições discretas e contínuas. Ela afirma que, para qualquer número k que satisfaça $k \geq 1$, $P(|X - \mu| \geq k\sigma) \leq 1/k^2$ (consulte o Exercício 43 no Capítulo 3 para obter uma interpretação). Obtenha essa probabilidade no caso de uma distribuição normal para $k = 1, 2$ e 3 e a compare ao limite superior.

48. Seja por X o número de falhas em um rolo de 100 m de fita magnética (uma variável de valores inteiros). Suponha que X tenha uma distribuição aproximadamente normal com $\mu = 25$ e $\sigma = 5$. Use a correção de continuidade para calcular a probabilidade de o número de falhas:
a. estar entre 20 e 30, inclusive.
b. ser no máximo 30. Inferior a 30.

49. Assuma X com uma distribuição binomial de parâmetros $n = 25$ e p. Calcule cada uma das probabilidades a seguir, usando a aproximação normal (com correção de continuidade) para os casos $p = 0{,}5$, $0{,}6$ e $0{,}8$ e compare com as probabilidades exatas calculadas pela Tabela A.1 do Apêndice.
a. $P(15 \leq X \leq 20)$
b. $P(X \leq 15)$
c. $P(20 \leq X)$

50. Suponha que 10% de todos os eixos de aço produzidos por um processo específico não estejam em conformidade mas possam ser retrabalhados (em vez de serem sucateados). Considere uma amostra aleatória de 200 eixos e represente por X o número dentre eles que não estejam conforme e possam ser retrabalhados. Qual é a probabilidade (aproximada) de X ser:
a. No máximo 30?

b. Menos que 30?

c. Estar entre 15 e 25 (inclusive)?

51. Suponha que apenas 40% de todos os motoristas de um estado específico usem o cinto de segurança regularmente. É selecionada uma amostra de 500 motoristas. Qual é a probabilidade de:

a. entre 180 e 230 (inclusive) motoristas usarem o cinto de segurança regularmente?

b. menos de 175 dos motoristas da amostra usarem o cinto regularmente? Menos de 150?

52. Mostre que a relação entre um percentil normal geral e o percentil correspondente z atende ao que foi dito nesta seção.

53. a. Mostre que, se X tem distribuição normal com parâmetros μ e σ, então $Y = aX + b$ (função linear de X) também tem distribuição normal. Quais são os parâmetros da distribuição de Y [ou seja, $E(Y)$ e $V(Y)$]? [*Sugestão*: escreva a FDA de Y, $P(Y \leq y)$ como uma integral que envolve a fdp de X e então faça a diferenciação em relação a y para obter a fdp de Y.]

b. Se, quando medida em °C, a temperatura tem distribuição normal com média 115 e desvio padrão 2, o que se pode dizer sobre a distribuição da temperatura medida em °F?

54. Não há uma boa fórmula para a fdc normal padrão $\Phi(z)$, mas foram publicadas diversas aproximações razoáveis. A aproximação a seguir foi obtida em "Approximations for Hand Calculators Using Small Integer Coefficients" (*Mathematics of Computation*, 1977, p. 214-222). Para $0 \leq z \leq 5{,}5$,

$$P(Z \geq z) = 1 - \Phi(z)$$

$$\approx 0{,}5 \exp\left\{-\left[\frac{(83z + 351)z + 562}{703/z + 165}\right]\right\}$$

O erro relativo dessa aproximação é inferior a 0,042%. Use isso para calcular aproximações das probabilidades a seguir e as compare com as probabilidades obtidas na Tabela A.3 do Apêndice, sempre que possível.

a. $P(Z \geq 1)$

b. $P(Z < -3)$

c. $P(-4 < Z < 4)$

d. $P(Z > 5)$

4.4 A Distribuição Gama e Seus Parentes

O gráfico de qualquer fdp normal tem formato de sino e é simétrico. Há várias situações práticas em que a variável de interesse do investigador pode ter distribuição com inclinação. Uma família de fdp que fornece diversos formatos de distribuições com inclinação é a família gama. Para defini-la, precisamos antes apresentar uma função que tem importante papel em vários ramos da matemática.

DEFINIÇÃO

Para $\alpha > 0$, **a função gama** $\Gamma(\alpha)$ é definida por

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx \qquad (4.5)$$

As propriedades mais importantes da função gama são as seguintes:

1. Para qualquer $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$ [via integração por partes]
2. Para qualquer inteiro positivo, n, $\Gamma(n) = (n - 1)!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Pela Expressão (4.6), se tivermos

$$f(x; \alpha) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & \text{caso contrário} \end{cases} \qquad (4.6)$$

então $f(x; \alpha) \geq 0$ e $\int_0^\infty f(x; \alpha)\,dx = \Gamma(\alpha)/\Gamma(\alpha) = 1$, então $f(x; \alpha)$ satisfaz às duas propriedades básicas de uma fdp.

A Família de Distribuições Gama

DEFINIÇÃO

Diz-se que uma variável aleatória contínua X tem uma **distribuição gama** se a fdp de X é

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & \text{caso contrário} \end{cases} \qquad (4.7)$$

onde os parâmetros α e β satisfazem $\alpha > 0$, $\beta > 0$. A **distribuição gama padrão** tem $\beta = 1$, de forma que a fdp de uma va gama padrão é dada por (4.6).

A Figura 4.26(a) ilustra os gráficos da fdp gama $f(x; \alpha, \beta)$ para diversos pares (α, β), enquanto a Figura 4.26(b) apresenta os gráficos da fdp gama padrão. Para a fdp-padrão, quando $\alpha \leq 1$, $f(x; \alpha)$ for estritamente decrescente à media que x aumenta a partir de 0. Quando $\alpha > 1$, $f(x; \alpha)$ aumenta a partir de 0 em $x = 0$ até um máximo e depois decresce. O parâmetro β em (4.7) é denominado *parâmetro de escala* porque os valores diferentes de 1 esticam ou comprimem a fdp na direção de x.

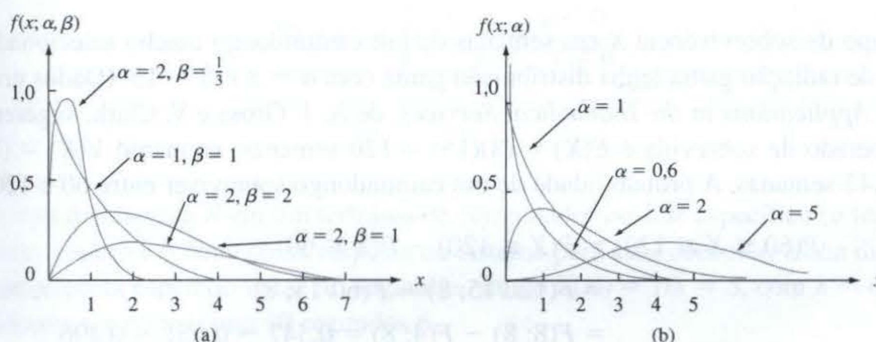


Figura 4.26 (a) Curvas de densidade gama; (b) curvas de densidade gama padrão

$E(X)$ e $E(X^2)$ podem ser obtidos a partir de uma integração razoavelmente direta e então $V(X) = E(X^2) - [E(X)]^2$.

A média e a variância de uma variável aleatória X com distribuição gama $f(x; \alpha, \beta)$ são

$$E(X) = \mu = \alpha\beta \qquad V(X) = \sigma^2 = \alpha\beta^2$$

Quando X é uma va gama-padrão, a fdc de X,

$$F(x; \alpha) = \int_0^x \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)} dy \qquad x > 0 \qquad (4.8)$$

é denominada **função gama incompleta** [algumas vezes a função gama incompleta se refere à Expressão (4.8) sem o denominador $\Gamma(\alpha)$ no integrando]. Há extensas tabelas de $F(x; \alpha)$ disponíveis. Na Tabela A.4 do Apêndice, apresentamos uma pequena tabulação para $\alpha = 1,2,..., 10$ e $x = 1,2,..., 15$.

Exemplo 4.20

Suponha que o tempo de reação X de um indivíduo selecionado aleatoriamente a um certo estímulo possui uma distribuição gama padrão com $\alpha = 2$ segundos. Como

$$P(a \leq X \leq b) = F(b) - F(a)$$

quando X é contínuo,

$$P(3 \le X \le 5) = F(5; 2) - F(3; 2) = 0{,}960 - 0{,}801 = 0{,}159$$

A probabilidade de o tempo de reação ser mais de 4 segundos é

$$P(X > 4) = 1 - P(X \le 4) = 1 - F(4; 2) = 1 - 0{,}908 = 0{,}092$$ ■

A função gama incompleta também pode ser usada para calcular probabilidades que envolvam distribuições gama não-padrão.

PROPOSIÇÃO Seja X com uma distribuição gama de parâmetros α e β. Então, para qualquer $x > 0$, a fdc de X é dada por

$$P(X \le x) = F(x; \alpha, \beta) = F\left(\frac{x}{\beta}; \alpha\right)$$

onde $F(\cdot; \alpha)$ é a função gama incompleta.[2]

Exemplo 4.21

Suponha que o tempo de sobrevivência X em semanas de um camundongo macho selecionado aleatoriamente exposto a 240 rads de radiação gama tenha distribuição gama com $\alpha = 8$ e $\beta = 15$. (Dados em *Survival Distributions: Reliability Applications in the Biomedical Services*, de A. J. Gross e V. Clark, sugerem $\alpha \approx 8{,}5$ e $\beta \approx 13{,}3$.) O tempo esperado de sobrevida é $E(X) = (8)(15) = 120$ semanas, enquanto $V(X) = (8)(15)^2 = 1800$ e $\sigma_X = \sqrt{1800} = 42{,}43$ semanas. A probabilidade de um camundongo sobreviver entre 60 e 120 semanas é

$$\begin{aligned} P(60 \le X \le 120) &= P(X \le 120) - P(X \le 60) \\ &= F(120/15; 8) - F(60/15; 8) \\ &= F(8; 8) - F(4; 8) = 0{,}547 - 0{,}051 = 0{,}496 \end{aligned}$$

A probabilidade de um camundongo sobreviver ao menos 30 semanas é

$$\begin{aligned} P(X \ge 30) &= 1 - P(X < 30) = 1 - P(X \le 30) \\ &= 1 - F(30/15; 8) = 0{,}999 \end{aligned}$$ ■

A Distribuição Exponencial

A família de distribuições exponenciais fornece modelos probabilísticos largamente usados em engenharia e em várias disciplinas da ciência.

DEFINIÇÃO Diz-se que X tem uma **distribuição exponencial** com parâmetro λ ($\lambda > 0$) se a fdp de X for

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & \text{caso contrário} \end{cases} \tag{4.9}$$

A fdp exponencial é um caso especial da Expressão (4.7) da fdp gama geral em que $\alpha = 1$ e β foi substituído por $1/\lambda$ [alguns autores usam a forma $(1/\beta)e^{-x/\beta}$]. A média e a variância de X são

$$\mu = \alpha\beta = \frac{1}{\lambda} \qquad \sigma^2 = \alpha\beta^2 = \frac{1}{\lambda^2}$$

[2] MINITAB e outros programas estatísticos calculam $F(x; \alpha, \beta)$, uma vez que os valores de x, α, e β são especificados.

Tanto a média como o desvio padrão da distribuição exponencial são iguais a $1/\lambda$. São exibidos diversos gráficos das fdps exponenciais na Figura 4.27.

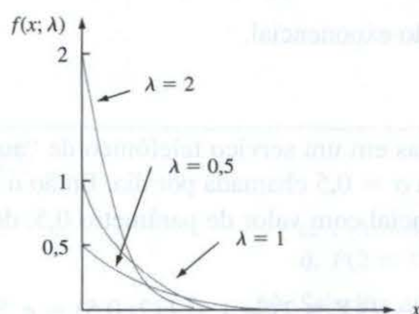


Figura 4.27 Curvas de densidade exponencial

Diferentemente da fdp gama geral, a fdp exponencial pode ser facilmente integrada. Em particular, a fdc de X é

$$F(x;\lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Exemplo 4.22

Suponha que o tempo de resposta X em um terminal de computador *on-line* específico (o tempo entre o final de uma consulta de um usuário e o começo da resposta do sistema para essa consulta) tenha distribuição exponencial com tempo de resposta esperado igual a 5 segundos. Então $E(X) = 1/\lambda = 5$, com $\lambda = 0{,}2$. A probabilidade de o tempo de resposta ser no máximo 10 segundos é

$$P(X \leq 10) = F(10; 0{,}2) = 1 - e^{-(0{,}2)(10)} = 1 - e^{-2} = 1 - 0{,}135 = 0{,}865$$

A probabilidade de o tempo de resposta estar entre 5 e 10 segundos é

$$P(5 \leq X \leq 10) = F(10; 0{,}2) - F(5; 0{,}2)$$
$$= (1 - e^{-2}) - (1 - e^{-1}) = 0{,}233 \quad ■$$

A distribuição exponencial é usada freqüentemente como modelo para distribuição dos tempos entre a ocorrência de eventos sucessivos, tais como clientes chegando em uma unidade de atendimento ou chamadas em uma central telefônica. O motivo disso é que a distribuição exponencial está fortemente relacionada ao processo de Poisson discutido no Capítulo 3.

PROPOSIÇÃO

Suponha que o número de eventos que ocorrem em um intervalo de tempo de duração t tenha distribuição de Poisson com parâmetro αt (onde α, a taxa do processo do evento, é o número esperado de eventos que ocorrem em uma unidade de tempo) e que os números das ocorrências em intervalos não-sobrepostos sejam independentes um do outro. Então a distribuição do tempo decorrido entre a ocorrência de dois eventos sucessivos é exponencial com parâmetro $\lambda = \alpha$.

Apesar de uma demonstração completa estar além do escopo do livro, o resultado é facilmente verificado para o tempo X_1, até que ocorra o primeiro evento:

$$P(X_1 \leq t) = 1 - P(X_1 > t) = 1 - P[\text{nenhum evento em } (0, t)]$$
$$= 1 - \frac{e^{-\alpha t} \cdot (\alpha t)^0}{0!} = 1 - e^{-\alpha t}$$

que é exatamente a FDA da distribuição exponencial.

Exemplo 4.23

Suponha que sejam recebidas chamadas em um serviço telefônico de "auxílio a suicidas" 24 horas, de acordo com um processo de Poisson com taxa $\alpha = 0{,}5$ chamada por dia. Então o número de dias X entre chamadas sucessivas tem uma distribuição exponencial com valor de parâmetro 0,5, de forma que a probabilidade de haver mais de 2 dias entre chamadas é

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F(2; 0{,}5) = e^{-(0{,}5)(2)} = 0{,}368$$

O tempo esperado entre chamadas sucessivas é $1/0{,}5 = 2$ dias. ■

Outra aplicação importante da distribuição exponencial é modelar a distribuição do tempo de vida de um componente. O motivo parcial da popularidade dessas aplicações é a propriedade de "falta de memória" da distribuição exponencial. Suponha que o tempo de vida de um componente seja distribuído exponencialmente com o parâmetro λ. Após o componente ser colocado em serviço, saímos por um período de t_0 horas e então retornamos para encontrar o componente ainda funcionando. Nesse caso, qual é a probabilidade de ele durar ao menos t horas adicionais? Em símbolos, queremos $P(X \geq t + t_0 \mid X \geq t_0)$. Pela definição de probabilidade condicional,

$$P(X \geq t + t_0 \mid X \geq t_0) = \frac{P[(X \geq t + t_0) \cap (X \geq t_0)]}{P(X \geq t_0)}$$

Entretanto, o evento, $X \geq t_0$ no numerador é redundante, já que os dois eventos podem ocorrer se, e somente se, $X \geq t + t_0$. Portanto,

$$P(X \geq t + t_0 \mid X \geq t_0) = \frac{P(X \geq t + t_0)}{P(X \geq t_0)} = \frac{1 - F(t + t_0; \lambda)}{1 - F(t_0; \lambda)} = e^{-\lambda t}$$

Tal probabilidade condicional é idêntica à probabilidade original $P(X \geq t)$ de o componente ter durado t horas. Assim, *a distribuição do tempo de vida adicional é exatamente igual à distribuição original de tempo de vida*, de forma que em cada ponto no tempo o componente não dá sinais de desgaste. Em outras palavras, a distribuição do tempo de vida restante é independente da idade atual.

Embora a propriedade de ausência de memória possa ser justificada pelo menos, aproximadamente, em muitos problemas aplicados, em outras situações os componentes se deterioram ou ocasionalmente melhoram com o uso (ao menos até certo ponto). Modelos mais gerais de tempo de vida são fornecidos pelas distribuições gama, Weibull e lognormal (as duas últimas serão discutidas na próxima seção).

A Distribuição Qui-Quadrado

DEFINIÇÃO

Seja ν um inteiro positivo. Diz-se que uma variável aleatória X possui uma **distribuição qui-quadrado** com parâmetro ν se a fdp de X for a densidade gama com $\alpha = \nu/2$ e $\beta = 2$. A fdp de uma va qui-quadrado será

$$f(x; \nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (4.10)$$

O parâmetro ν é denominado **número de graus de liberdade** (gl) de X. O símbolo χ^2 freqüentemente é usado no lugar de "qui-quadrado".

A distribuição qui-quadrado é importante por ser a base de diversos procedimentos de inferência estatística. O motivo disso é que as distribuições qui-quadrado estão intimamente relacionadas a distribuições normal (consulte o Exercício 65). Discutiremos a distribuição qui-quadrado com mais detalhes nos capítulos sobre inferência.

Exercícios | Seção 4.4 (55–65)

55. Calcule os dados a seguir:
a. $\Gamma(6)$
b. $\Gamma(5/2)$
c. $F(4; 5)$ (a função gama incompleta)
d. $F(5; 4)$
e. $F(0; 4)$

56. Seja X uma distribuição gama-padrão com $\alpha = 7$. Calcule o seguinte:
a. $P(X \leq 5)$
b. $P(X < 5)$
c. $P(X > 8)$
d. $P(3 \leq X \leq 8)$
e. $P(3 < X < 8)$
f. $P(X < 4$ ou $X > 6)$

57. Suponha que o tempo gasto por um aluno selecionado aleatoriamente que usa um terminal conectado a uma instalação de computador com *time-sharing* tem uma distribuição gama com média de 20 minutos e variância de 80 min².
a. Quais são os valores de α e β?
b. Qual é a probabilidade de um aluno usar o terminal por no máximo 24 minutos?
c. Qual é a probabilidade de um aluno passar entre 20 e 40 minutos usando o terminal?

58. Suponha que, quando um certo tipo de transistor é sujeito a um teste de vida acelerado, o tempo de vida X (em semanas) possui distribuição gama com média 24 semanas e desvio padrão 12 semanas.
a. Qual é a probabilidade de um transistor durar entre 12 e 24 semanas?
b. Qual é a probabilidade de um transistor durar no máximo 24 semanas? A mediana da distribuição do tempo de vida é inferior a 24? Por que ou por que não?
c. Qual é o 99º percentil da distribuição do tempo de vida?
d. Suponha que o teste seja finalizado após t semanas. Que valor de t faz com que apenas 0,5% de todos os transistores estejam operando no fim do teste?

59. Seja X = tempo entre duas chegadas sucessivas no guichê de atendimento rápido de um banco local. Se X possui distribuição exponencial com $\lambda = 1$ (que é idêntico a uma distribuição gama-padrão com $\alpha = 1$), calcule os itens a seguir:
a. O tempo esperado entre duas chegadas sucessivas
b. O desvio padrão do tempo entre chegadas sucessivas
c. $P(X \leq 4)$
d. $P(2 \leq X \leq 5)$

60. Seja X a distância (m) que um animal viaja desde seu local de nascimento até o primeiro local vago que encontra. Suponha que, para ratos-cangurus de rabo de bandeira, X possui uma distribuição exponencial com parâmetro $\lambda = 0{,}01386$ (como sugerido no artigo "Competition and Dispersal from Multiple Nests," *Ecology*, 1997, p. 873-883).
a. Qual é a probabilidade de a distância ser no máximo 100 m? No máximo 200 m? Entre 100 e 200 m?
b. Qual é a probabilidade de a distância exceder a média por mais de 2 desvios padrão?
c. Qual é o valor da mediana da distância?

61. Diversas experiências com determinado tipo de ventilador, usados em motores a diesel, indicam que a distribuição exponencial sugere um bom modelo para cálculo do tempo até uma falha. Suponha que o tempo médio seja 25.000 horas. Qual é a probabilidade de:
a. um ventilador selecionado aleatoriamente durar pelo menos 20.000 horas? No máximo 30.000 horas? Entre 20.000 e 30.000 horas?
b. o tempo de vida de um ventilador exceder o valor médio em mais de 2 desvios padrão? Mais de 3 desvios padrão?

62. O caso especial da distribuição gama em que α é um inteiro positivo n é denominado distribuição de Erlang. Se substituirmos β por $1/\lambda$ na Expressão (4.7), a fdp Erlang será

$$f(x; \lambda, n) = \begin{cases} \dfrac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

É possível demonstrar que, se os tempos entre os eventos sucessivos forem independentes, cada um com uma distribuição exponencial de parâmetro λ, o tempo total X antes de todos os n próximos eventos ocorrerem terá fdp $f(x; \lambda, n)$.
a. Qual é o valor esperado de X? Se o tempo (em minutos) entre as chegadas de clientes sucessivos for distribuído exponencialmente com $\lambda = 0{,}5$, quanto tempo podemos esperar que passe antes da chegada do décimo cliente?
b. Se o tempo entre as chegadas de clientes for distribuído exponencialmente com $\lambda = 0{,}5$, qual é a probabilidade de o décimo cliente (depois do que

acabou de chegar) chegue dentro dos próximos 30 minutos?

c. O evento $\{X \leq t\}$ ocorre se e somente se ao menos n eventos ocorrerem nas próximas t unidades de tempo. Use o fato de que o número de eventos que ocorrem em um intervalo de duração t possui uma distribuição de Poisson com parâmetro λt para escrever uma expressão (envolvendo probabilidades de Poisson) para a FDA de Erlang $F(t; \lambda, n) = P(X \leq t)$.

63. Um sistema consiste de cinco componentes idênticos conectados em série, conforme ilustrado:



Se um componente falhar, todo o sistema falhará. Suponha que cada componente tenha um tempo de vida distribuído exponencialmente com $\lambda = 0{,}01$ e que eles falhem de forma independente. Defina os eventos A_i = {i-ésimo componente que dura no mínimo t horas }, $i = 1, \ldots, 5$, de forma que os A_is são eventos independentes. Seja X = momento da falha do sistema, isto é, o tempo de vida mais curto (mínimo) entre os cinco componentes.

a. O evento $\{X \geq t\}$ é equivalente a que evento que envolva $A_1, \ldots, A_5$?

b. Usando a independência dos A_is, calcule $P(X \geq t)$. Obtenha então $F(t) = P(X \leq t)$ e a fdp de X. Que tipo de distribuição tem X?

c. Suponha que haja n componentes, cada um com tempo de vida exponencial com parâmetro λ. Que tipo de distribuição tem X?

64. Se X tiver distribuição exponencial com parâmetro λ, deduza uma expressão geral para o percentil $(100p)$ da distribuição. Então faça a especialização para obter a mediana.

65. a. O evento $\{X^2 \leq y\}$ é equivalente a que evento que envolve X?

b. Se X tiver distribuição normal padrão, use a parte (a) para escrever a integral que iguala $P(X^2 \leq y)$. Diferencie então essa expressão em relação a y para obter a fdp de X^2 [o quadrado de uma variável $N(0, 1)$]. Finalmente, mostre que X^2 possui distribuição qui-quadrado com $\nu = 1$ gl [consulte (4.10)]. (*Sugestão:* use a identidade a seguir).

$$\frac{d}{dy}\left\{\int_{a(y)}^{b(y)} f(x)\,dx\right\} = f[b(y)] \cdot b'(y) - f[a(y)] \cdot a'(y)$$

4.5 Outras Distribuições Contínuas

As famílias de distribuições normal, gama (incluindo exponencial) e uniforme fornecem ampla variedade de modelos probabilísticos para variáveis contínuas, mas há muitas situações práticas em que nenhum membro dessas famílias consegue se enquadrar bem em um conjunto de dados observados. Os estatísticos e outros investigadores desenvolveram outras famílias de distribuições que normalmente são apropriadas, na prática.

A Distribuição de Weibull

A família de distribuições de Weibull foi apresentada pelo físico sueco Waloddi Weibull, em 1939. Seu artigo de 1951 "A Statistical Distribution Function of Wide Applicability" (*J. Applied Mechanics*, vol. 18, p. 293-297) apresenta diversas aplicações.

DEFINIÇÃO Uma variável aleatória X tem distribuição de Weibull com parâmetros α e β ($\alpha > 0, \beta > 0$) se a fdp de X for

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{4.11}$$

Em algumas situações, há justificativas teóricas para a utilização da distribuição de Weibull, mas em muitas aplicações $f(x; \alpha, \beta)$ simplesmente fornece uma boa aproximação dos dados observados para valores específicos de α e β. Quando $\alpha = 1$, a fdp é reduzida à distribuição exponencial (com $\lambda = 1/\beta$), de forma que a distribuição

exponencial é um caso especial das distribuições gama e de Weibull. Entretanto, há distribuições gama que não são distribuições de Weibull e vice-versa, o que faz com que uma família não seja um subconjunto da outra. Tanto α como β podem variar para se obter diversos formatos de distribuição diferentes, como ilustrado na Figura 4.28. β é um parâmetro de escala, de forma que valores diferentes esticam ou comprimem o gráfico na direção de x.

Integrar para obter $E(X)$ e $E(X^2)$ resulta em

$$\mu = \beta\Gamma\left(1 + \frac{1}{\alpha}\right) \qquad \sigma^2 = \beta^2\left\{\Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2\right\}$$

O cálculo de μ e de σ^2 necessita, portanto, da função gama.

A integração $\int_0^x f(y;\ \alpha, \beta)\, dy$ pode ser feita facilmente para obter a fdc de X.

A fdc de uma va de Weibull com parâmetros α e β é

$$F(x;\alpha,\beta) = \begin{cases} 0 & x < 0 \\ 1 - e^{-(x/\beta)^{\alpha}} & x \geq 0 \end{cases} \tag{4.12}$$

Exemplo 4.24

Nos últimos anos, a distribuição de Weibull tem sido usada para modelar emissões de poluentes de vários motores. Seja X o valor da emissão de NO_x (g/gal) a partir de certo tipo de motor de quatro tempos selecionado aleatoriamente e suponha que X possua uma distribuição de Weibull com $\alpha = 2$ e $\beta = 10$ (sugerido pelas informações do artigo "Quantification of Variability and Uncertainty in Lawn and Garden Equipment NO_x and Total Hydrocarbon Emission Factors," *J. of the Air and Waste Management Assoc.*, 2002, p. 435-448). A curva de densidade correspondente tem a mesma aparência daquela da Figura 4.28 para $\alpha = 2$, $\beta = 1$, exceto pelo fato de que agora os valores 50 e 100 substituem 5 e 10 no eixo horizontal (porque β é uma "parâmetro de escala"). Então

$$P(X \leq 10) = F(10; 2, 10) = 1 - e^{-(10/10)^2} = 1 - e^{-1} = 0{,}632$$

De forma similar, $P(X \leq 25) = 0{,}998$, de maneira que a distribuição está quase inteiramente concentrada nos valores entre 0 e 25. O valor c, que separa os 5% de todos os motores que emitem as maiores quantidades de NO_x dos 95% restantes, satisfaz

$$0{,}95 = 1 - e^{-(c/10)^2}$$

Isolando o termo exponencial em um lado, tirando os logaritmos e resolvendo a equação resultante, temos $c \approx 17{,}3$ como o 95º percentil da distribuição de tempos de vida. ■

Freqüentemente, em situações práticas, o modelo de Weibull é razoável, exceto pelo fato de que o menor valor X possível pode ser um valor γ não assumido como zero (isso também se aplica a um modelo gama). A quantidade γ pode então ser vista como um terceiro parâmetro da distribuição, que foi o que Weibull fez em seu trabalho original. Para, digamos, $\gamma = 3$, todas as curvas na Figura 4.28 seriam deslocadas 3 unidades para a direita. Isso é equivalente a dizer que $X - \gamma$ possui a fdp (4.11), de forma que a fdc de X é obtida pela substituição de x em (4.12) por $x - \gamma$.

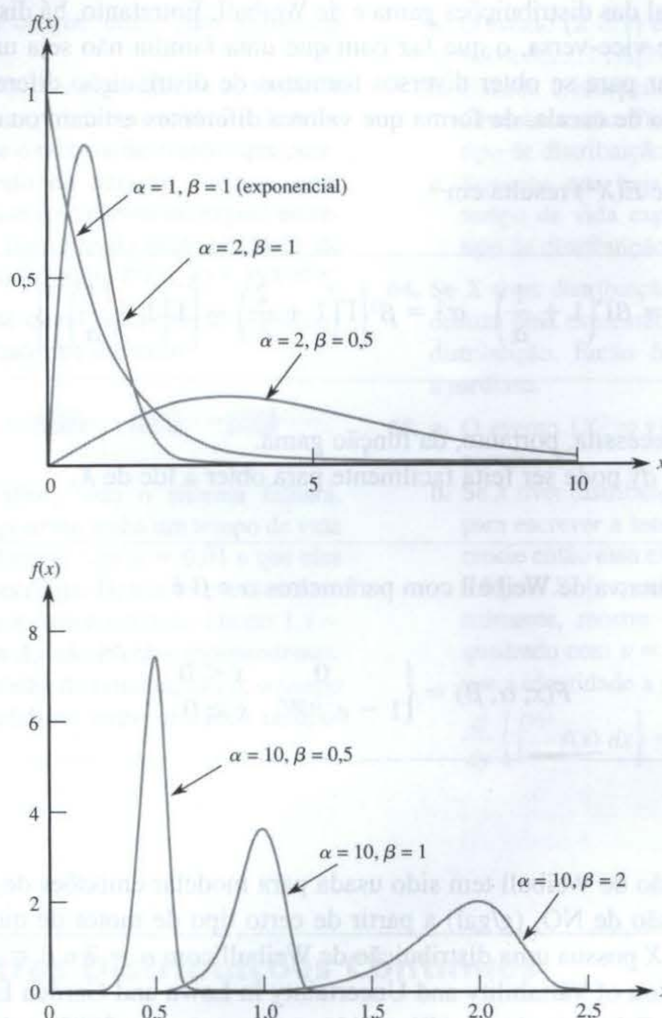


Figura 4.28 Curvas de densidade de Weibull

Exemplo 4.25

Seja X = perda de massa por corrosão de um pequeno prato quadrado de uma liga de magnésio imersa durante 7 dias em uma solução aquosa inibida de 20% de $MgBr_2$. Suponha que a menor perda de massa possível seja $\gamma = 3$ e que o excesso $X - 3$, além desse mínimo, tenha distribuição de Weibull com $\alpha = 2$ e $\beta = 4$. (Esse exemplo foi considerado em "Practical Applications of the Weibull Distribution," *Industrial Quality Control*, Ago., 1964, p. 71-78. Os valores de α e β foram escolhidos como 1,8 e 3,67, respectivamente, embora no artigo tenha sido usada uma escolha de parâmetros ligeiramente diferente.) A fdc de X será então

$$F(x; \alpha, \beta, \gamma) = F(x; 2, 4, 3) = \begin{cases} 0 & x < 3 \\ 1 - e^{-[(x-3)/4]^2} & x \geq 3 \end{cases}$$

Portanto,

$$P(X > 3{,}5) = 1 - F(3{,}5; 2, 4, 3) = e^{-0{,}0156} = 0{,}985$$

e

$$P(7 \leq X \leq 9) = 1 - e^{-2{,}25} - (1 - e^{-1}) = 0{,}895 - 0{,}632 = 0{,}263$$ ■

A Distribuição Lognormal

DEFINIÇÃO

Diz-se que uma va não-negativa X possui uma **distribuição lognormal** se a va $Y = \ln(X)$ possui uma distribuição normal. A fdp resultante de uma va lognormal quando $\ln(X)$ tiver distribuição normal com parâmetros μ e σ é

$$f(x; \mu, \sigma) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\,\sigma x} e^{-[\ln(x)-\mu]^2/(2\sigma^2)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Tenha cuidado neste caso. Os parâmetros μ e σ não são a média e o desvio padrão de X e sim de $\ln(X)$. A média e a variância de X são

$$E(X) = e^{\mu+\sigma^2/2} \qquad V(X) = e^{2\mu+\sigma^2} \cdot (e^{\sigma^2} - 1)$$

No Capítulo 5, apresentaremos uma justificativa teórica para essa distribuição, relacionada ao Teorema do Limite Central, mas, assim como ocorre com outras distribuições, a lognormal pode ser usada como modelo, mesmo na ausência de tal justificativa. A Figura 4.29 ilustra os gráficos da fdp lognormal. Embora a curva normal seja simétrica, a lognormal possui inclinação positiva.

Como $\ln(X)$ possui distribuição normal, a fdc de X pode ser expressa em termos da fdc $\Phi(z)$ de uma va Z normal padrão. Para $x \geq 0$,

$$\begin{aligned} F(x; \mu, \sigma) &= P(X \leq x) = P[\ln(X) \leq \ln(x)] \\ &= P\left(Z \leq \frac{\ln(x) - \mu}{\sigma}\right) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) \end{aligned} \qquad (4.13)$$

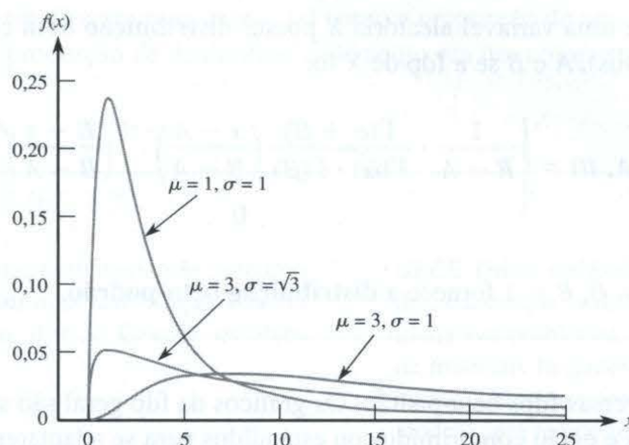


Figura 4.29 Curvas de densidade lognormal

Exemplo 4.26

A distribuição lognormal é usada freqüentemente como modelo para diversas propriedades de materiais. O artigo "Reliability of Wood Joist Floor Systems with Creep" (*J. of Structural Engr.*, 1995, p. 946-954) sugere que a distribuição lognormal com $\mu = 0{,}375$ e $\sigma = 0{,}25$ é um modelo plausível para X = módulo da elasticidade

(MOE, em 10^6 psi) de sistemas de vigas de madeira construídos de hem-fir grau número 2. O valor da média e a variância de MOE são

$$E(X) = e^{0,375 + (0,25)^2/2} = e^{0,40625} = 1,50$$
$$V(X) = e^{0,8125}(e^{0,0625} - 1) = 0,1453$$

A probabilidade de MOE estar entre 1 e 2 é

$$\begin{aligned} P(1 \le X \le 2) &= P(\ln(1) \le \ln(X) \le \ln(2)) \\ &= P(0 \le \ln(X) \le 0,693) \\ &= P\left(\frac{0 - 0,375}{0,25} \le Z \le \frac{0,693 - 0,375}{0,25}\right) \\ &= \Phi(1,27) - \Phi(-1,50) = 0,8312 \end{aligned}$$

Que valor c faz com que apenas 1% dos sistemas tenha MOE que exceda c? Desejamos o c para que

$$0,99 = P(X \le c) = P\left(Z \le \frac{\ln(c) - 0,375}{0,25}\right)$$

a partir de onde $(\ln(c) - 0,375)/0,25 = 2,33$ e $c = 2,605$. Portanto, 2,605 é o 99º percentil da distribuição de MOE. ■

Distribuição Beta

Todas as famílias de distribuições contínuas discutidas até agora, exceto a distribuição uniforme, têm densidade positiva em um intervalo infinito (apesar de normalmente a função densidade decrescer rapidamente para zero além de alguns desvios padrão em relação à média). A distribuição beta fornece densidade positiva apenas para X em um intervalo de comprimento finito.

DEFINIÇÃO

Diz-se que uma variável aleatória X possui **distribuição beta** com parâmetros α, β (ambos positivos), A e B se a fdp de X for

$$f(x;\alpha,\beta,A,B) = \begin{cases} \dfrac{1}{B-A}\cdot\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\left(\dfrac{x-A}{B-A}\right)^{\alpha-1}\left(\dfrac{B-x}{B-A}\right)^{\beta-1} & A \le x \le B \\ 0 & \text{caso contrário} \end{cases}$$

O caso $A = 0$, $B = 1$ fornece a **distribuição beta-padrão.**

A Figura 4.30 ilustra diversas fdps beta-padrão. Os gráficos da fdp geral são similares, exceto pelo fato de que os valores são deslocados e então comprimidos ou estendidos para se adaptarem a $[A, B]$. A menos que α e β sejam inteiros, a integração da fdp para calcular probabilidades é difícil, de forma que normalmente é usada uma tabela da função beta incompleta. A média e a variância de X são

$$\mu = A + (B - A)\cdot\frac{\alpha}{\alpha+\beta} \qquad \sigma^2 = \frac{(B-A)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

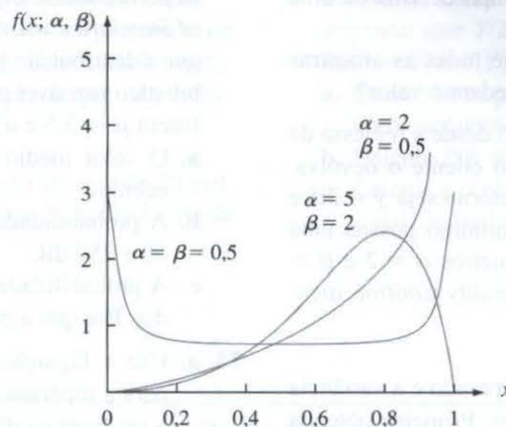


Figura 4.30 Curvas de densidade beta-padrão

Exemplo 4.27

Os gerentes de projeto normalmente usam um método denominado PERT (técnica de revisão e avaliação de programa) para coordenar as diversas atividades de um projeto grande (uma aplicação de sucesso foi a construção da espaçonave Apollo.) Uma hipótese-padrão na análise PERT é que o tempo necessário para completar qualquer atividade específica depois de seu início tem uma distribuição beta com A = tempo otimista (se tudo correr bem) e B = tempo pessimista (se tudo correr mal). Suponha que, na construção de uma casa, o tempo X (em dias) necessário para a construção das fundações possui distribuição beta com $A = 2$, $B = 5$, $\alpha = 2$, e $\beta = 3$. Então $\alpha/(\alpha + \beta) = 0,4$, de forma que $E(X) = 2 + (3)(0,4) = 3,2$. Para esses valores de a e β, a fdp de X é uma função polinomial simples. A probabilidade de levar no máximo 3 dias para construir os alicerces é

$$P(X \leq 3) = \int_2^3 \frac{1}{3} \cdot \frac{4!}{1!2!} \left(\frac{x-2}{3}\right)\left(\frac{5-x}{3}\right)^2 dx$$

$$= \frac{4}{27}\int_2^3 (x-2)(5-x)^2\, dx = \frac{4}{27} \cdot \frac{11}{4} = \frac{11}{27} = 0,407$$ ■

A distribuição beta-padrão normalmente é usada para modelar a variação na proporção ou porcentagem de uma quantidade que ocorre em diferentes amostras, tal como a proporção de um dia de 24 horas em que um indivíduo está dormindo ou a proporção de determinado elemento em um composto químico.

Exercícios Seção 4.5 (66–80)

66. O tempo de vida X (em centenas de horas) de certo tipo de válvula eletrônica possui uma distribuição Weibull com parâmetros $\alpha = 2$ e $\beta = 3$. Calcule os itens a seguir:
a. $E(X)$ e $V(X)$
b. $P(X \leq 6)$
c. $P(1,5 \leq X \leq 6)$
(Essa distribuição de Weibull é sugerida como modelo de tempo em serviço em "On the Assessment of Equipment Reliability: Trading Data Collection Costs for Precision," *J. Engr Manuf.*, 1991, p. 105-109.)

67. Os autores do artigo "A Probabilistic Insulation Life Model for Combined Thermal-Electrical Stresses" (*IEEE Trans. on Elect. Insulation*, 1985, p. 519-522) declararam que "a distribuição de Weibull é largamente usada em problemas estatísticos relativos ao desgaste de materiais isolantes sólidos sujeitos ao desgaste e à tensão". Eles propõem o uso da distribuição como um modelo para tempo (em horas) de falha de amostras de isolantes sólidos sujeitos a voltagem de AC. Os valores dos parâmetros dependem da voltagem e da temperatura. Suponha que $\alpha = 2,5$ e $\beta = 200$ (valores sugeridos pelos dados no artigo)
a. Qual é a probabilidade de o tempo de vida de um espécime ser no máximo 250? Inferior a 250? Maior de 300?

b. Qual é a probabilidade de o tempo de vida de uma amostra estar entre 100 e 250?

c. Que valor faz com que 50% de todas as amostras tenham tempos de vida que excedam o valor?

68. Seja X = tempo (em 10^{-1} semanas) desde a remessa de um produto com defeito até que o cliente o devolva. Suponha que o tempo mínimo de retorno seja $\gamma = 3{,}5$ e que o excesso $X - 3{,}5$ acima do mínimo possua uma distribuição de Weibull com parâmetros $\alpha = 2$ e $\beta = 1{,}5$ (consulte o artigo *Industrial Quality Control*, mencionado no Exemplo 4.25).

a. Qual é a FDA de X?

b. Quais são o tempo de retorno esperado e a variância do tempo de retorno? [*Sugestão*: Primeiro obtenha $E(X - 3{,}5)$ e $V(X - 3{,}5)$.]

c. Calcule $P(X > 5)$.

d. Calcule $P(5 \leq X \leq 8)$.

69. Seja X com distribuição de Weibull e a fdp da Expressão (4.11). Demonstre que $\mu = \beta\Gamma(1 + 1/\alpha)$. (*Sugestão*: na integral de $E(X)$, faça a mudança de variável $y = (x/\beta)^{\alpha}$, de forma que $x = \beta y^{1/\alpha}$.)

70. a. No Exercício 66, qual a mediana do tempo de vida das válvulas? [*Sugestão*: use a Expressão (4.12).]

b. No Exercício 68, qual é a mediana do prazo de retorno?

c. Se X tiver uma distribuição de Weibull com a fdc da Expressão (4.12), obtenha uma expressão geral para o ($100p$)ésimo percentil da distribuição.

d. No Exercício 68, a empresa deseja recusar devoluções após t semanas. Para que valores de t apenas 10% de todas as devoluções serão recusadas?

71. Seja X o limite de resistência à tração (ksi) a −200° de um corpo de prova de aço selecionado aleatoriamente que apresenta "fragilidade ao frio" em baixas temperaturas. Suponha que X tenha distribuição de Weibull com $\alpha = 20$ e $\beta = 100$.

a. Qual é a probabilidade de X ser no máximo 105?

b. Se forem selecionados vários corpos de prova um após o outro, qual é a probabilidade no longo prazo de obter valores de resistência entre 100 e 105 ksi?

c. Qual é o valor da mediana da distribuição de resistência?

72. Os autores do artigo em que foram obtidos os dados do Exercício 1.27 sugerem que um modelo de probabilidade razoável para a vida útil de uma broca é a distribuição lognormal com $\mu = 4{,}5$ e $\sigma = 0{,}8$.

a. Qual é o valor da média e o desvio padrão da vida útil?

b. Qual é a probabilidade da vida útil ser no máximo 100?

c. Qual é a probabilidade da vida útil ser no mínimo 200? Maior que 200?

73. Seja X = potência média horária (em decibéis) de sinais de rádio recebidos transmitidos entre duas cidades. Os autores do artigo "Families of Distributions for Hourly Median Power and Instantaneous Power of Received Radio Signals" (*J. Research National Bureau of Standards*, vol. 67D, 1963, p. 753-762) argumentam que a distribuição lognormal fornece um modelo probabilístico razoável para X. Se os valores dos parâmetros forem $\mu = 3{,}5$ e $\sigma = 1{,}2$, calcule os itens a seguir:

a. O valor médio e o desvio-padrão da potência recebida.

b. A probabilidade de a potência recebida estar entre 50 e 250 dB.

c. A probabilidade de X ser menor que o valor da média. Por que a probabilidade não é 0,5?

74. a. Use a Equação (4.13) para escrever uma fórmula para a mediana $\tilde{\mu}$ da distribuição lognormal. Qual é a mediana da distribuição do Exercício 73?

b. Lembrando que z_α é a notação do percentil $100(1 - \alpha)$ da distribuição padrão normal, escreva uma expressão para o percentil $100(1 - \alpha)$ da distribuição lognormal. No Exercício 73, que valor a potência recebida excederá apenas 5% do tempo?

75. Uma justificativa teórica baseada no mecanismo de falha de certo material admite a hipótese de que a resistência dúctil X de um material possui distribuição lognormal. Suponha que os parâmetros sejam $\mu = 5$ e $\sigma = 0{,}1$.

a. Calcule $E(X)$ e $V(X)$.

b. Calcule $P(X > 125)$.

c. Calcule $P(110 \leq X \leq 125)$.

d. Qual é o valor da mediana da resistência dúctil?

e. Se 10 diferentes amostras de uma liga de aço desse tipo forem submetidas a um teste de resistência, quantas você espera que tenham ao menos 125?

f. Se 5% dos menores valores não forem aceitáveis, qual seria a resistência mínima aceitável?

76. O artigo "The Statistics of Phytotoxic Air Pollutants" (*J. Royal Stat. Soc.*, 1989, p. 183-198) sugere a distribuição lognormal com um modelo de concentração de SO_2 sobre uma floresta específica. Suponha que os valores dos parâmetros sejam $\mu = 1{,}9$ e $\sigma = 0{,}9$.

a. Qual é o valor médio e o desvio padrão da concentração?

b. Qual é a probabilidade de a concentração ser no máximo 10? Entre 5 e 10?

77. Que condição de α e β é necessária para que a fdp beta-padrão seja simétrica?

78. Suponha que a proporção X da área coberta por uma determinada planta em um quadrante selecionado aleatoriamente possui distribuição beta-padrão com $\alpha = 5$ e $\beta = 2$.

a. Calcule $E(X)$ e $V(X)$.

b. Calcule $P(X \leq 0{,}2)$.

c. Calcule $P(0{,}2 \leq X \leq 0{,}4)$.

d. Qual é a proporção esperada da região de amostragem não coberta pela planta?

79. Assuma que X tenha densidade beta-padrão com parâmetros α e β.

a. Demonstre a fórmula de $E(X)$ fornecida na seção.
b. Calcule $E[(1 - X)^m]$. Se X representa a proporção de uma substância formada por um determinado ingrediente, qual é a proporção esperada que não é formada por ele?

80. É aplicada uma força em uma barra de aço de 20 pol. presa nas duas extremidades. Seja Y = distância da quebra da barra em relação à extremidade esquerda. Suponha que $Y/20$ possui uma distribuição beta com $E(Y) = 10$ e $V(Y) = \frac{100}{7}$.
a. Quais são os parâmetros da distribuição beta-padrão correspondente?
b. Calcule $P(8 \leq Y \leq 12)$.
c. Calcule a probabilidade de a barra quebrar a mais de 2 pol. de onde você espera.

4.6 Gráficos de Probabilidade

Um investigador normalmente obtém uma amostra numérica $x_1, x_2, \ldots, x_n$ e deseja saber se é plausível ela ter vindo de uma distribuição de população de um tipo específico (por exemplo, de uma distribuição normal). Muitos procedimentos formais de inferência estatística se baseiam na hipótese de que a distribuição da população é de um tipo especificado. O uso de tal procedimento não é apropriado se a distribuição de probabilidades subjacente for muito diferente do tipo assumido. Além disso, a compreensão da distribuição subjacente algumas vezes pode fornecer percepções dos mecanismos físicos envolvidos na geração dos dados. Uma forma eficiente de verificar uma hipótese de distribuição é construir o que chamamos de **gráfico de probabilidade**. A essência desse gráfico é que, se a distribuição na qual ele se baseia estiver correta, seus pontos estarão próximos a uma linha reta. Se a distribuição real for muito diferente da usada para criar o gráfico, os pontos diferem substancialmente do padrão linear.

Percentis da Amostra

Os detalhes envolvidos na construção de gráficos de probabilidade são um pouco diferentes de fonte para fonte. A base da nossa construção é a comparação entre os percentis dos dados da amostra e os percentis correspondentes da distribuição consideradas. Lembre-se que o $(100p)$ésimo percentil de uma distribuição contínua com fdc $F(\cdot)$ é o número $\eta(p)$ que satisfaz $F(\eta(p)) = p$. Isto é, $\eta(p)$ é o número na escala de medida tal que a área sob a curva de densidade à esquerda de $\eta(p)$ é p. Assim, o 50º percentil $\eta(0{,}5)$ satisfaz $F(\eta(0{,}5)) = 0{,}5$, e o 90º percentil satisfaz $F(\eta(0{,}9)) = 0{,}9$. Considere como exemplo a distribuição-padrão normal, para a qual representamos a fdc por $\Phi(\cdot)$. Na Tabela A.3 do Apêndice, encontramos o 20º percentil localizando a linha e a coluna em que 0,2000 (ou um número o mais próximo dele) aparece na tabela. Como 0,2005 aparece na interseção da linha −0,8 e da coluna 0,04, o 20º percentil é aproximadamente −0,84. De forma similar, o 25º percentil da distribuição normal-padrão é (usando interpolação linear) aproximadamente −0,675.

Grosso modo, os percentis da amostra são definidos da mesma forma que os da distribuição da população. O 50º percentil da amostra deve separar os 50% menores valores dos 50% maiores da amostra, o 90º percentil deve ser o limite entre 90% abaixo do valor e 10% acima, e assim por diante. Infelizmente encontramos problemas quando tentamos calcular os percentis da amostra para uma amostra determinada de n observações. Se, por exemplo, $n = 10$, podemos dividir 20% desses valores ou 30% dos dados, mas nenhum valor dividirá exatamente 23% das 10 observações. Para nos aprofundarmos, precisamos de uma definição operacional de percentis da amostra (local em que diferentes pessoas fazem coisas ligeiramente diferentes). Lembre-se de que, quando n é ímpar, a mediana da amostra, ou o 50º percentil, é o valor do meio na lista ordenada, por exemplo, o sexto maior valor quando $n = 11$. Isto significa considerar a observação do centro localizando-se metade na parte inferior dos dados e metade na superior. De forma similar, suponha $n = 10$. Então, se denominarmos o terceiro menor valor como 25º percentil, estaremos considerando metade desse valor o grupo inferior (formado pelas duas menores observações) e metade no grupo superior (as sete maiores observações). Tal consideração leva à seguinte definição geral de percentis da amostra:

DEFINIÇÃO Ordene as n observações da amostra da menor para a maior. A menor observação da lista é então definida como o **[100(*i* − 0,5)/*n*]º percentil da amostra.**

Uma vez calculados os valores percentuais de $100(i - 0{,}5)/n$ ($i = 1, 2, \ldots, n$), os percentis da amostra correspondentes aos percentuais intermediários são obtidos por interpolação linear. Por exemplo: se $n = 10$, os percentuais correspondentes às observações ordenadas serão $100(1 - 0{,}5)/10 = 5\%$, $100(2 - 0{,}5)/10 = 15\%$, $25\%, \ldots$, e $100(10 - 0{,}5)/10 = 95\%$. O 10º percentil estará a meio caminho entre o 5º (menor observação da amostra) e o 15º percentil (segunda menor observação da amostra). Para os nossos propósitos, tal interpolação não será necessária porque o gráfico de probabilidade se baseará somente nos percentuais $100(i - 0{,}5)/n$ correspondentes às n observações da amostra.

Um Gráfico de Probabilidade

Suponha agora que para os percentuais $100(i - 0{,}5)/n$ ($i = 1, \ldots, n$) os percentis sejam determinados para uma distribuição de população específica cuja plausibilidade está sendo investigada. Se a amostra foi realmente selecionada na distribuição especificada, os percentis da amostra (observações ordenadas da amostra) devem estar relativamente perto dos percentis correspondentes da distribuição da população. Isso é, para $i = 1, 2, \ldots, n$ deve haver concordância razoável entre a i-ésima menor observação da amostra e o $[100(i - 0{,}5)/n]$ésimo percentil da distribuição especificada. Considere os pares (percentil da população, percentil da amostra), isso é, os pares

$$\left(\begin{matrix}[100(i - 0{,}5)/n]\text{ésimo percentil} \\ \text{da distribuição}\end{matrix}, \begin{matrix} i\text{-ésima menor observação} \\ \text{da amostra}\end{matrix} \right)$$

para $i = 1, \ldots, n$. Cada par desse tipo é locado como um ponto em um sistema de coordenadas bidimensional. Se os percentis da amostra estiverem próximos aos percentis correspondentes da distribuição da população, o primeiro número de cada par será aproximadamente igual ao segundo. Os pontos do gráfico estão próximos a uma reta de 45°. Desvios substanciais dos pontos plotados em relação à reta de 45° geram dúvidas sobre a hipótese de a distribuição considerada ser a correta.

Exemplo 4.28

O valor de uma constante física é conhecido de um experimentador. Ele executa $n = 10$ medidas independentes desse valor, usando certo dispositivo de medida, e registra os erros de medida resultantes (erro = valor observado − valor real). Essas observações são exibidas na tabela a seguir.

Percentagem	5	15	25	35	45
percentil z	−1,645	−1,037	−0,675	−0,385	−0,126
Observação da amostra	−1,91	−1,25	−0,75	−0,53	0,20

Percentagem	55	65	75	85	95
percentil z	0,126	0,385	0,675	1,037	1,645
Observação da amostra	0,35	0,72	0,87	1,40	1,56

É plausível que a variável aleatória *erro de medida* tenha uma distribuição normal padrão? Os percentis (z) normais padronizados necessários também são exibidos na tabela. Portanto, os pontos no gráfico de probabilidade são (−1,645; −1,91), (−1,037; −1,25), ... e (1,645; 1,56). A Figura 4.31 (página 191) mostra o gráfico resultante. Apesar de os pontos se desviarem um pouco da reta de 45°, a impressão predominante é que a reta ajusta muito bem os pontos. O gráfico sugere que a distribuição normal padrão seja um modelo probabilístico razoável para o erro de medida.

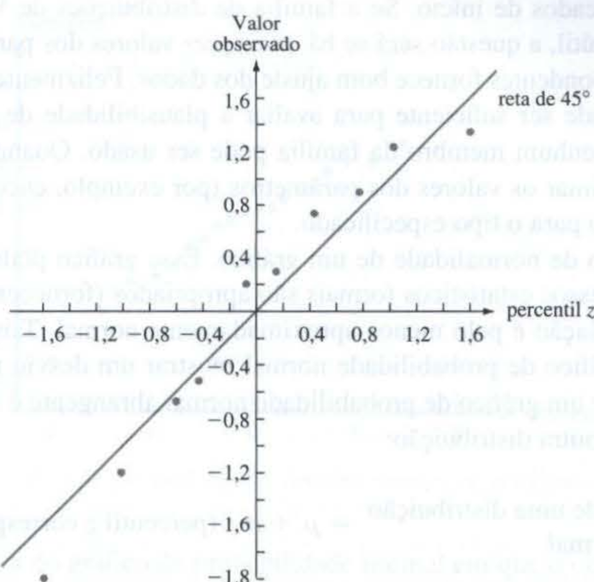


Figura 4.31 Gráfico de pares (percentil z, valor observado) dos dados do Exemplo 4.28: primeira amostra

A Figura 4.32 mostra um gráfico de pares (percentil z, observação) de uma segunda amostra de 10 observações. A reta de 45° fornece bom ajuste para a parte do meio da amostra, mas não para os extremos. O gráfico tem uma aparência bem-definida de formato S.

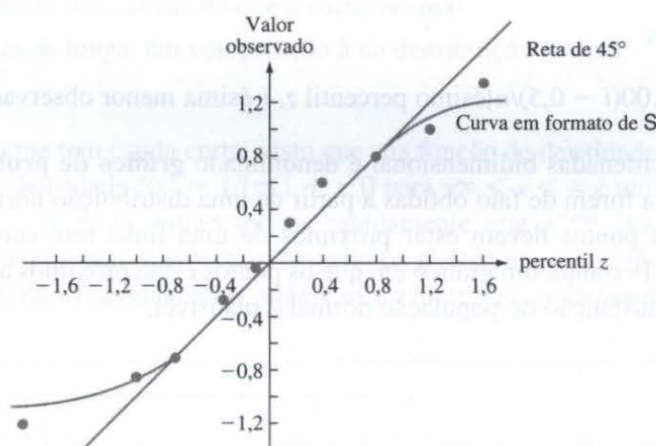


Figura 4.32 Gráfico de pares (percentil z, valor observado) dos dados do Exemplo 4.28: segunda amostra

As duas menores observações da amostra são consideravelmente maiores que os percentis z correspondentes (os pontos na extremidade esquerda do gráfico estão muito acima da reta de 45°). De forma similar, as duas maiores observações são muito menores do que os percentis z associados. Esse gráfico indica que a distribuição normal padronizada não seria uma escolha plausível para o modelo de probabilidade que propiciou o aparecimento desses erros de medida observados. ■

Normalmente um investigador não está interessado em saber apenas se uma distribuição probabilística especificada, tal como a distribuição normal padrão (normal com $\mu = 0$ e $\sigma = 1$) ou a distribuição exponencial com $\lambda = 0,1$, é um modelo plausível para a distribuição da população da qual a amostra foi selecionada. Em vez disso, o investigador vai querer saber se algum membro de uma família de distribuições de probabilidade especifica um modelo plausível — a família de distribuições normais, a família de distribuições exponenciais, a família de distribuições de Weibull e assim por diante. Os valores dos parâmetros de uma distribuição

normalmente não são especificados de início. Se a família de distribuições de Weibull for considerada como modelo para os dados de vida útil, a questão será se há *quaisquer* valores dos parâmetros α e β para os quais a distribuição de Weibull correspondentes fornece bom ajuste dos dados. Felizmente, quase sempre ocorre de apenas um gráfico de probabilidade ser suficiente para avaliar a plausibilidade de uma família. Se o gráfico se desviar demais de uma reta, nenhum membro da família pode ser usado. Quando o gráfico é bastante reto, é exigido mais trabalho para estimar os valores dos parâmetros (por exemplo, encontrar valores para μ e σ) que forneçam a melhor distribuição para o tipo especificado.

Enfoquemos a verificação de normalidade de um gráfico. Esse gráfico pode ser bastante útil no trabalho aplicado porque diversos processos estatísticos formais são apropriados (fornecem inferências precisas) apenas quando a distribuição da população é pelo menos aproximadamente normal. Tais procedimentos normalmente não devem ser usados se o gráfico de probabilidade normal mostrar um desvio muito pronunciado da linearidade. A chave para se construir um gráfico de probabilidade normal abrangente é a relação entre os percentis (z) normais padronizados e os de outra distribuição:

$$\begin{array}{l}\text{percentil de uma distribuição} \\ (\mu, \sigma) \text{ normal}\end{array} = \mu + \sigma \cdot (\text{percentil } z \text{ correspondente})$$

Considere primeiro o caso $\mu = 0$. Então, se cada observação for exatamente igual ao percentil normal correspondente para um valor particular de σ, os pares ($\sigma \cdot$ [percentil z], observação) estarão sobre uma reta de 45°, que tem coeficiente angular 1. Tal fato implica que os pares (percentil z, observação) localizam-se em uma reta que passa por (0, 0) (isto é, com intercepto y em 0) mas com coeficiente angular σ em vez de 1. O efeito de um valor de μ diferente de zero é simplesmente mudar o intercepto y de 0 para μ.

Um gráfico dos n pares

$$([100(i - 0{,}5)/n]\text{ésimo percentil } z,\ i\text{-ésima menor observação})$$

em um sistema de coordenadas bidimensional é denominado **gráfico de probabilidade normal**. Se as observações da amostra forem de fato obtidas a partir de uma distribuição normal com valor da média μ e desvio padrão σ, os pontos devem estar próximos de uma linha reta com coeficiente angular σ e cruzando o eixo em μ. Portanto, um gráfico em que os pontos estão próximos a uma linha reta sugere que a suposição de uma distribuição de população normal é plausível.

Exemplo 4.29

A amostra a seguir, formada por $n = 20$ observações sobre a voltagem de corte dielétrica de uma resina epóxi foi exibida no artigo "Maximum Likelihood Estimation in the 3-Parameter Weibull Distribution (*IEEE Trans. on Dielectrics and Elec. Insul.*, 1996, p. 43-55). Os valores de $(i - 0{,}5)/n$ para os quais os percentis z são obrigatórios são $(1 - 0{,}5)/20 = 0{,}025$, $(2 - 0{,}5)/20 = 0{,}075$,..., e 0,975.

Observação	24,46	25,61	26,25	26,42	26,66	27,15	27,31	27,54	27,74	27,94
Percentil z	−1,96	−1,44	−1,15	−0,93	−0,76	−0,60	−0,45	−0,32	−0,19	−0,06
Observação	27,98	28,04	28,28	28,49	28,50	28,87	29,11	29,13	29,50	30,88
Percentil z	0,06	0,19	0,32	0,45	0,60	0,76	0,93	1,15	1,44	1,96

A Figura 4.33 mostra o gráfico de probabilidade normal resultante. O padrão no gráfico é bastante reto, indicando que é plausível a distribuição da população de voltagens de corte dielétrica ser normal.

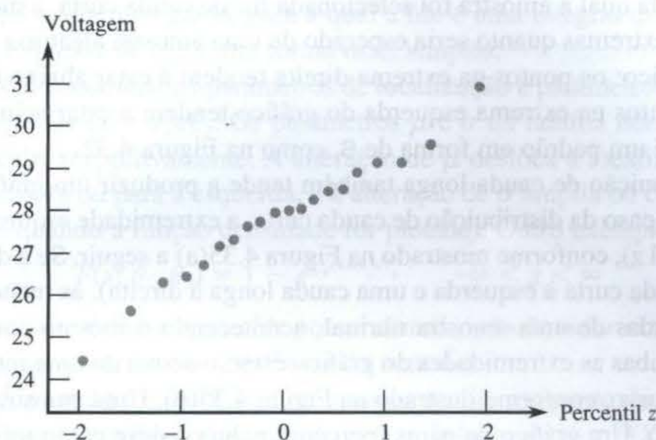


Figura 4.33 Gráfico de probabilidade normal da amostra de voltagem de dielétrica de corte ■

Há uma versão alternativa do gráfico de probabilidade normal em que o eixo dos percentis z é substituído por um eixo de probabilidade não-linear. A escala dos eixos é construída de forma que os pontos devem novamente estar próximos a uma reta quando a distribuição da amostra for normal. A Figura 4.34 a seguir mostra um gráfico do tipo em MINITAB, para os dados de voltagem de corte do Exemplo 4.29.

Uma distribuição de população não-normal freqüentemente pode se enquadrar em uma das três categorias a seguir:

1. É simétrica e tem "caudas mais curtas" do que uma distribuição normal, isto é, a curva de densidade decresce mais rapidamente nas caudas do que a curva normal.
2. É simétrica e com cauda longa, em comparação à da distribuição normal.
3. É inclinada.

Uma distribuição uniforme tem cauda curta, visto que sua função de densidade decresce a zero fora de um intervalo finito. A função de densidade $f(x) = 1/[\pi(1 + x^2)]$ para $-\infty < x < \infty$ é um exemplo de distribuição de cauda longa, pois $1/(1 + x^2)$ decresce muito menos rapidamente que $e^{-x^2/2}$. As distribuições lognormal e de Weibull estão entre as que possuem inclinação. Quando os pontos em um gráfico de probabilidade normal não tendem para a reta, o padrão freqüentemente sugere que a distribuição da população esteja em uma dessas três categorias.

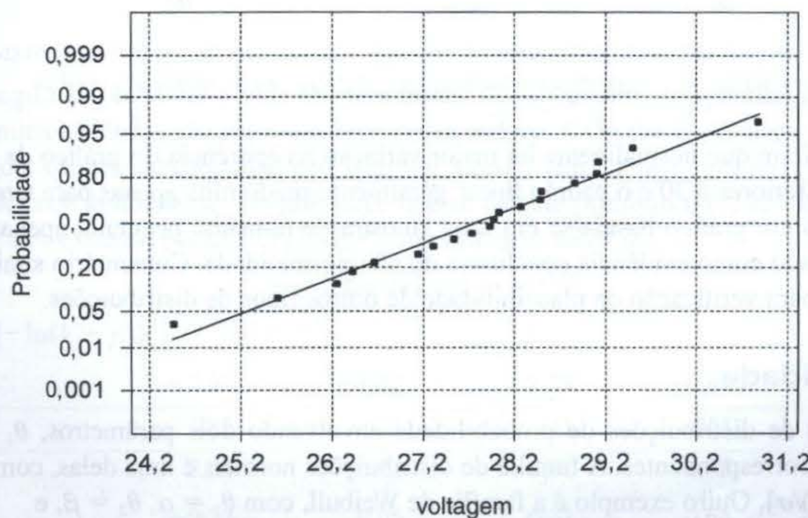


Figura 4.34 Gráfico de probabilidade normal dos dados de voltagem de corte em MINITAB

Quando a distribuição da qual a amostra foi selecionada for de cauda curta, a menor e a maior observações normalmente não serão tão extremas quanto seria esperado de uma amostra aleatória normal. Visualize uma reta desenhada no centro do gráfico: os pontos na extrema direita tendem a estar abaixo da reta (valor observado < percentil z), enquanto os pontos na extrema esquerda do gráfico tendem a estar acima da reta (valor observado > percentil z). O resultado é um padrão em forma de S, como na Figura 4.32.

Uma amostra de distribuição de cauda longa também tende a produzir um gráfico com formato de S. Entretanto, em contraste com o caso da distribuição de cauda curta, a extremidade esquerda do gráfico se curva para baixo (observado < percentil z), conforme mostrado na Figura 4.35(a) a seguir. Se a distribuição subjacente tiver inclinação positiva (uma cauda curta à esquerda e uma cauda longa à direita), as menores observações amostrais serão maiores que as esperadas de uma amostra normal, acontecendo o mesmo com as maiores observações. Nesse caso, os pontos em ambas as extremidades do gráfico estarão acima de uma reta que passa pela parte central, resultando um padrão curvo, conforme ilustrado na Figura 4.35(b). Uma amostra de distribuição lognormal geralmente produz tal padrão. Um gráfico de pares (percentil z, $\ln(x)$) deve então ter a aparência de uma reta.

Mesmo quando a distribuição da população é normal, os percentis da amostra não coincidirão exatamente com os percentis teóricos por causa da variabilidade da amostragem. Quanto os pontos do gráfico de probabilidade podem se desviar do padrão da reta antes que a hipótese de normalidade da população não seja mais plausível? A resposta a essa questão não é fácil. *Grosso modo*, uma pequena amostra de distribuição normal tem maior probabilidade de gerar um gráfico com padrão não-linear do que uma amostra grande. O livro *Fitting Equations to Data* (veja a bibliografia do Capítulo 13) apresenta os resultados de um estudo de simulação em que diversas amostras de diferentes tamanhos foram selecionadas de distribuições normais.

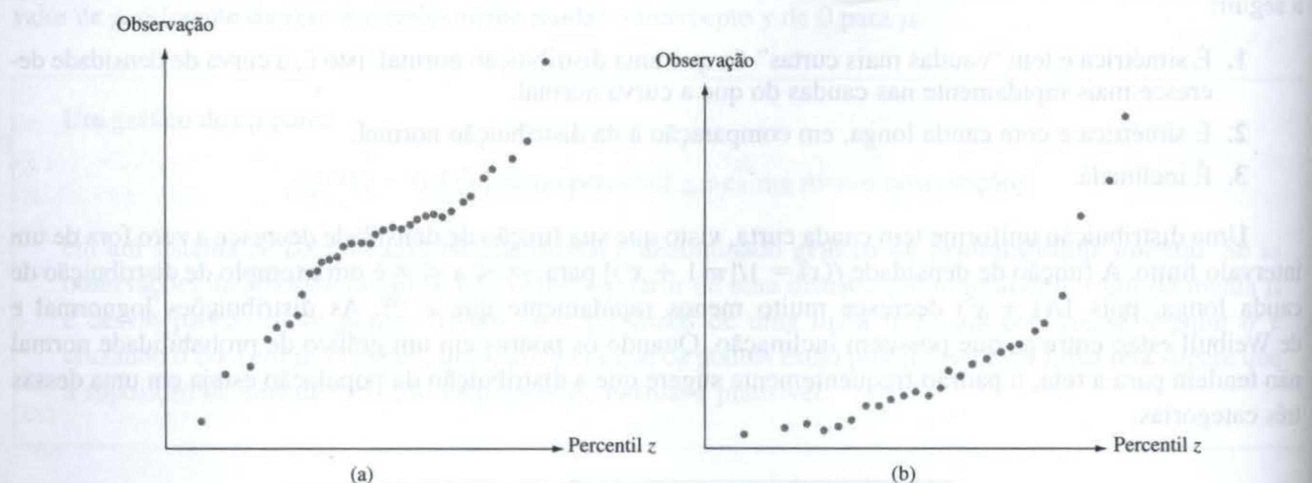


Figura 4.35 Gráficos de probabilidade que sugerem distribuição não-normal: (a) um gráfico consistente com distribuição de cauda longa; (b) um gráfico consistente com distribuição de inclinação positiva

Os autores concluíram que normalmente há maior variação na aparência do gráfico de probabilidade para tamanhos de amostra inferiores a 30 e o padrão linear geralmente predomina apenas para tamanhos de amostras muito maiores. Quando um gráfico baseia-se em uma amostra de tamanho pequeno, apenas um desvio muito substancial pode ser usado como evidência conclusiva de não-normalidade. Comentário similar se aplica a gráficos de probabilidade para verificação de plausibilidade de outros tipos de distribuições.

Além da Normalidade

Considere uma família de distribuições de probabilidade envolvendo dois parâmetros, θ_1 e θ_2, e represente por $F(x; \theta_1, \theta_2)$ as fdcs correspondentes. A família de distribuições normais é uma delas, com $\theta_1 = \mu$, $\theta_2 = \sigma$, e $F(x; \mu, \sigma) = \Phi[(x - \mu)/\sigma]$. Outro exemplo é a família de Weibull, com $\theta_1 = \alpha$, $\theta_2 = \beta$, e

$$F(x; \alpha, \beta) = 1 - e^{-(x/\beta)^\alpha}$$

Outra família desse tipo é a família gama, para a qual a fdc é uma integral que envolve a função gama incompleta que não pode ser expressa de nenhuma forma mais simples.

Os parâmetros θ_1 e θ_2 são denominados parâmetros de **localização** e parâmetros de **escala**, respectivamente, se $F(x; \theta_1, \theta_2)$ for uma função de $(x - \theta_1)/\theta_2$. Os parâmetros μ e σ da família normal são parâmetros de localização e parâmetros de escala, respectivamente. A alteração de μ desloca a localização da curva de densidade em formato de sino para a direita ou para a esquerda, e a alteração de σ amplia ou comprime a escala de medida (a escala do eixo horizontal quando a função densidade for plotada). Outro exemplo é dado pela fdc

$$F(x; \theta_1, \theta_2) = 1 - e^{-e^{(x-\theta_1)/\theta_2}} \qquad -\infty < x < \infty$$

Diz-se que uma variável aleatória com essa fdc possui *distribuição de valor extremo*. É usada em aplicações que envolvem a vida útil de componente e a resistência do material.

Apesar de a fdc do valor extremo sugerir à primeira vista que θ_1 é o ponto de simetria da função de densidade e, portanto, ser a média e a mediana, esse não é o caso. Pelo contrário, $P(X \leq \theta_1) = F(\theta_1; \theta_1, \theta_2) = 1 - e^{-1} = 0{,}632$, e a função de densidade $f(x; \theta_1, \theta_2) = F'(x; \theta_1, \theta_2)$ tem inclinação negativa (cauda longa mais baixa). De forma similar, o parâmetro da escala θ_2 não é o desvio padrão ($\mu = \theta_1 - 0{,}5772\theta_2$ e $\sigma = 1{,}283\theta_2$). Entretanto, a mudança do valor de θ_1 altera a localização da curva de densidade, enquanto a alteração de θ_2 redefine o eixo de medidas.

O parâmetro β da distribuição de Weibull é um parâmetro de escala, mas α não é de localização. O parâmetro α normalmente é mencionado como **parâmetro de formato**. Um comentário similar se aplica aos parâmetros α e β da distribuição gama. Na forma usual, a função de densidade de qualquer membro da distribuição gama ou de Weibull é positiva para $x \Rightarrow 0$ e zero em caso contrário. Um parâmetro de localização pode ser introduzido como um terceiro parâmetro γ (fizemos isso para a distribuição de Weibull) para deslocar a função de densidade de forma que seja positiva se $x \Rightarrow \gamma$ e zero em caso contrário.

Quando a família em consideração possui apenas parâmetros de escala e de localização, a questão de algum membro da família ser ou não uma distribuição de população plausível pode ser resolvida por meio de um único gráfico de probabilidade de fácil construção. Primeiro são obtidos os percentis da distribuição-padrão, aquela com $\theta_1 = 0$ e $\theta_2 = 1$, para porcentagens $100(i - 0{,}5)/n$ $(i = 1,..., n)$. Os n (percentil padronizado, observação) pares fornecem os pontos do gráfico. Isso foi exatamente o que fizemos para obter um gráfico de probabilidade normal abrangente. Surpreendentemente, essa metodologia pode ser aplicada para obtenção de um gráfico de probabilidade de Weibull. O resultado-chave é que, se X tiver distribuição de Weibull com parâmetro de formato α e parâmetro de escala β, a variável transformada $\ln(X)$ tem uma distribuição de valor extremo com parâmetro de localização $\theta_1 = \ln(\beta)$ e parâmetro de escala α. Assim, um gráfico dos pares (percentil do valor extremo padronizado, $\ln(x)$) que mostre padrão linear forte fornece suporte para a escolha da distribuição Weibull como modelo de população.

Exemplo 4.30

As observações a seguir são relativas à vida útil (em horas) do isolamento de aparelho elétrico quando a aceleração da tensão térmica e elétrica são mantidas com certos valores ("On the Estimation of Life of Power Apparatus Insulation Under Combined Electrical and Thermal Stress," *IEEE Trans. on Electrical Insulation*, 1985, p. 70-78). O gráfico de probabilidade de Weibull exige que se calcule primeiro o 5º, 15º, ésimo ... e 95º percentis da distribuição de valor extremo padrão. O $(100p)^{\text{o}}$ percentil $\eta(p)$ satisfaz

$$p = F(\eta(p)) = 1 - e^{-e^{\eta(p)}}$$

de onde $\eta(p) = \ln[-\ln(1 - p)]$.

Percentil	−2,97	−1,82	−1,25	−0,84	−0,51
x	282	501	741	851	1072
ln(x)	5,64	6,22	6,61	6,75	6,98

Percentil	−0,23	0,05	0,33	0,64	1,10
x	1122	1202	1585	1905	2138
ln(x)	7,02	7,09	7,37	7,55	7,67

Os pares (−2,97, 5,64), (−1,82, 6,22),..., (1,10, 7,67) estão plotados como pontos na Figura 4.36. A disposição dos pontos em reta é uma forte evidência de que se deve usar a distribuição de Weibull como modelo da vida útil do isolamento, conclusão a que o autor do artigo citado também chegou.

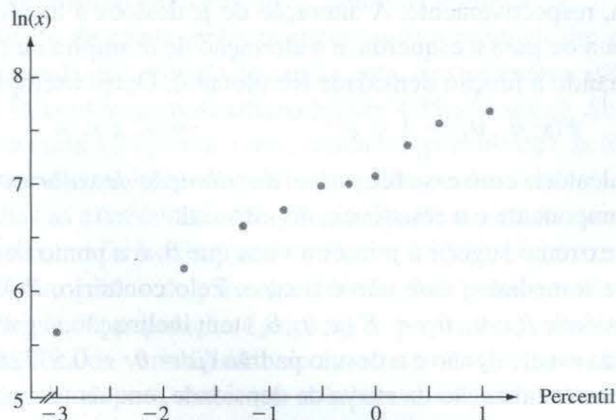
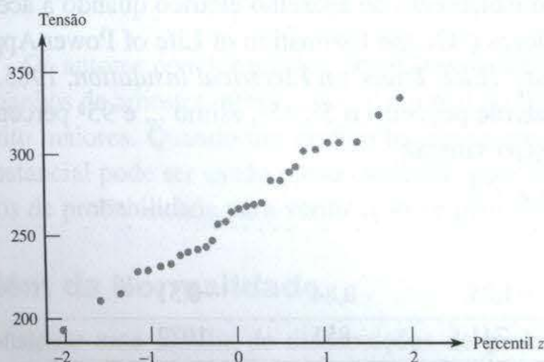


Figura 4.36 Gráfico de probabilidade de Weibull dos dados de vida útil do isolamento ■

A distribuição gama é um exemplo de uma família que envolve um parâmetro de formato para o qual não há transformação $h(\cdot)$, tal que $h(X)$ tenha distribuição que depende apenas dos parâmetros de localização e escala. A construção de um gráfico de probabilidade requer primeiro a estimativa do parâmetro de formato pelos dados da amostra (alguns métodos de processamento são descritos no Capítulo 6). Às vezes, um investigador quer saber se a variável transformada X^θ tem distribuição normal para algum valor de θ (por convenção, $\theta = 0$ é identificado com a transformação logarítmica, caso em que X possui distribuição lognormal). O livro *Graphical Methods for Data Analysis*, relacionado na bibliografia do Capítulo 1, aborda esse tipo de problema, bem como outras discussões dos gráficos de probabilidade.

Exercícios | Seção 4.6 (81–91)

81. O gráfico de probabilidade normal a seguir foi construído a partir de uma amostra de 30 leituras de tensão na malha de pontos de tubos de vídeo usados em monitores de computadores. Parece plausível que a distribuição da tensão seja normal?



82. Considere as 10 observações a seguir de vida útil de mancais (em horas):

152,7	172,0	172,5	173,3	193,0
204,7	216,5	234,9	262,6	422,6

Construa o gráfico de probabilidade normal e comente sobre a plausibilidade da distribuição normal como modelo de estimativa da vida útil de mancais (dados obtidos em "Modified Moment Estimation for the Three-Parameter Lognormal Distribution," *J. Quality Technology*, 1985, p. 92 -99).

83. Construa um gráfico de probabilidade normal para a seguinte amostra de observações relativas à espessura da cobertura de tintas de baixa viscosidade ("Achieving a Target Value for a Manufacturing Process: A Case Study", *J. of Quality Technology*, 1992, p. 22-26). Você se sentiria confortável ao estimar a média da espessura da população usando um método que assume uma distribuição de população normal?

0,83	0,88	0,88	1,04	1,09	1,12	1,29	1,31
1,48	1,49	1,59	1,62	1,65	1,71	1,76	1,83

84. O artigo "A Probabilistic Model of Fracture in Concrete and Size Effects on Fracture Toughness" (*Magazine of Concrete Res.*, 1996, p. 311-320) fornece argumentos que explicam por que a distribuição de dureza de fratura em corpos de prova de concreto tem uma distribuição de

Weibull, além de apresentar diversos histogramas de dados que parecem ajustados com curvas de Weibull superpostas. Considere a seguinte amostra de tamanho $n = 18$ observações de dureza de concreto de alta resistência (consistente com um dos histogramas). Os valores de $p_i = (i - 0{,}5)/18$ também são fornecidos.

Observação	0,47	0,58	0,65	0,69	0,72	0,74
p_i	0,0278	0,0833	0,1389	0,1944	0,2500	0,3056
Observação	0,77	0,79	0,80	0,81	0,82	0,84
p_i	0,3611	0,4167	0,4722	0,5278	0,5833	,6389
Observação	0,86	0,89	0,91	0,95	1,01	1,04
p_i	0,6944	0,7500	0,8056	0,8611	,09167	0,9722

Construa um gráfico de probabilidade de Weibull e comente.

85. Construa um gráfico de probabilidade normal para os dados de propagação de trincas por fadiga do Exercício 39 (Capítulo 1). Parece plausível que a vida da propagação tenha distribuição normal? Explique.

86. O artigo "The Load-Life Relationship for M50 Bearings with Silicon Nitride Ceramic Balls" (*Lubrication Engr.*, 1984, p. 153-159) relata os dados a seguir sobre a vida de carga de mancais (milhões revs.) para mancais testados com cargas de 6,45 kN.

47,1	68,1	68,1	90,8	103,6	106,0	115,0
126,0	146,6	229,0	240,0	240,0	278,0	278,0
289,0	289,0	367,0	385,9	392,0	505,0	

a. Construa um gráfico de probabilidade normal. A normalidade é plausível?
b. Construa um gráfico de probabilidade de Weibull. A família de distribuições Weibull é plausível?

87. Construa um gráfico de probabilidade que permita avaliar a normalidade da distribuição lognormal como um modelo para os dados de precipitação do Exercício 81 (Capítulo 1).

88. As observações a seguir são valores de precipitação no mês de março de um período de 30 anos em Minneapolis-St. Paul.

0,77	1,20	3,00	1,62	2,81	2,48
1,74	0,47	3,09	1,31	1,87	0,96
0,81	1,43	1,51	0,32	1,18	1,89
1,20	3,37	2,10	0,59	1,35	0,90
1,95	2,20	0,52	0,81	4,75	2,05

a. Construa e interprete um gráfico de distribuição normal para o conjunto de dados.
b. Calcule a raiz quadrada de cada valor e então construa um gráfico de probabilidade normal com base nos dados transformados. Parece plausível que a raiz quadrada da precipitação tenha distribuição normal?
c. Repita a parte (b) após transformar por raízes cúbicas.

89. Use um software estatístico para construir um gráfico de probabilidade normal dos dados do limite de resistência à tração fornecidos no Exercício 13 do Capítulo 1 e comente.

90. Sejam as observações ordenadas da amostra $y_1, y_2, \ldots, y_n$ (y_1 sendo o menor e y_n o maior). Sugerimos, para verificação de normalidade, que os pares $(\Phi^{-1}((i - 0{,}5)/n), y_i)$ sejam plotados. Suponha que acreditemos que as observações são provenientes de uma distribuição com média 0 e assuma $w_1, \ldots, w_n$ como os valores absolutos ordenados dos x_is. Um gráfico **seminormal** é um gráfico de probabilidade dos w_is. Mais especificamente, como $P(|Z| \leq w) = P(-w \leq Z \leq w) = 2\Phi(w) - 1$, um gráfico seminormal é um gráfico dos pares $(\Phi^{-1}((i - 0{,}5)/n) + 1)/2, w_i)$. A virtude desse gráfico é que os pequenos ou grandes *outliers* na amostra original serão exibidos apenas na extremidade superior e não nas duas extremidades. Construa um gráfico seminormal para a seguinte amostra de erros de medida e comente: −3,78, −1,27, 1,44, −0,39, 12,38, −43,40, 1,15, −3,96, −2,34, 30,84.

91. As seguintes observações de tempos de falha (milhares de horas) são resultantes de testes de vida aceleradas de 16 chipes de certo tipo de circuito integrado:

82,8	11,6	359,5	502,5	307,8	179,7
242,0	26,5	244,8	304,3	379,1	212,6
229,9	558,9	366,7	204,6		

Use os percentis correspondentes da distribuição exponencial com $\lambda = 1$ para construir um gráfico de probabilidade. Explique, então, por que o gráfico avalia a plausibilidade da amostra gerada a partir de qualquer distribuição exponencial.

Exercícios Suplementares (92–120)

92. Seja X = tempo que leva para uma cabeça de leitura/gravação encontrar um registro específico em uma memória em disco de um computador, após ter sido posicionada sobre a faixa correta. Se os discos executarem uma rotação em cada 25 milisegundos, uma hipótese razoável é que X seja distribuído uniformemente no intervalo [0, 25].

a. Calcule $P(10 \leq X \leq 20)$.
b. Calcule $P(X \geq 10)$.
c. Deduza a fdc $F(X)$.
d. Calcule $E(X)$ e σ_X.

93. Uma barra de 12 pol. fixa nas duas extremidades será submetida a uma tensão crescente até sua ruptura.

Seja Y = distância da ruptura da barra em relação à extremidade esquerda. Suponha que Y tenha fdp

$$f(y) = \begin{cases} \left(\frac{1}{24}\right)y\left(1 - \frac{y}{12}\right) & 0 \leq y \leq 12 \\ 0 & \text{caso contrário} \end{cases}$$

Calcule os dados a seguir:

a. A fdc de Y, e faça o gráfico.

b. $P(Y \leq 4)$, $P(Y > 6)$, e $P(4 \leq Y \leq 6)$

c. $E(Y)$, $E(Y^2)$ e $V(Y)$

d. A probabilidade de o ponto de ruptura ocorrer a mais de 2 pol. do ponto de ruptura esperado.

e. O comprimento esperado do menor segmento após a ruptura.

94. Seja X o tempo até a falha (em anos) de certo componente hidráulico. Suponha que a fdp de X seja $f(x) = 32/(x + 4)^3$ para $x > 0$.

a. Demonstre que se $f(x)$ é uma fdp legítima.

b. Determine a fdc.

c. Use o resultado da parte (b) para calcular a probabilidade de o tempo até a falha estar entre 2 e 5 anos.

d. Qual é o tempo esperado até a falha?

e. Se o componente tiver valor residual de $100/(4 + x)$, quando o tempo até a falha for x, qual será o valor residual esperado?

95. O tempo de conclusão X de uma determinada tarefa possui fdc $F(x)$ dada por

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^3}{3} & 0 \leq x < 1 \\ 1 - \frac{1}{2}\left(\frac{7}{3} - x\right)\left(\frac{7}{4} - \frac{3}{4}x\right) & 1 \leq x \leq \frac{7}{3} \\ 1 & x \geq \frac{7}{3} \end{cases}$$

a. Obtenha a fdp $f(x)$ e faça o gráfico.

b. Calcule $P(0{,}5 \leq X \leq 2)$.

c. Calcule $E(X)$.

96. A tensão de ruptura de um determinado tipo de diodo escolhido aleatoriamente possui distribuição normal com valor de média 40 V e desvio padrão 1,5 V.

a. Qual é a probabilidade de a voltagem de um diodo estar entre 39 e 42?

b. Que valor faz com que apenas 15% dos diodos tenham voltagens que excedam aquele valor?

c. Se quatro diodos forem selecionados de forma independente, qual será a probabilidade de ao menos um ter uma voltagem maior que 42?

97. O artigo "Computer Assisted Net Weight Control" (*Quality Progress*, 1983, p. 22-25) sugere uma distribuição normal com média 137,2 onças e desvio padrão 1,6 onça para o conteúdo real de um tipo específico de jarras. O conteúdo nominal é de 135 onças.

a. Qual é a probabilidade de uma única jarra conter mais do que o conteúdo nominal?

b. Entre 10 jarras selecionadas aleatoriamente, qual é a probabilidade de ao menos oito conterem mais do que o conteúdo nominal?

c. Assumindo que a média permaneça em 137,2, para que valor o desvio padrão teria de ser alterado para que 95% de todas as jarras contenham mais do que o conteúdo nominal?

98. Quando as placas de circuito usadas na fabricação de CD-players são testadas, a porcentagem de peças com defeito no longo prazo é 5%. Suponha que um lote de 250 placas tenha sido recebido e que a condição de uma placa é independente das outras.

a. Qual é a probabilidade aproximada de que ao menos 10% das placas do lote tenham defeito?

b. Qual é a probabilidade aproximada de que exatamente 10 peças tenham defeito no lote?

99. O artigo "Characterization of Room Temperature Damping in Aluminum-Indium Alloys" (*Metallurgical Trans.*, 1993, p. 1611-1619) sugere que o tamanho de grão de matriz Al (μm) de uma liga formada por 2% de índio pode ser modelado com uma distribuição normal de valor médio 96 e desvio padrão 14.

a. Qual é a probabilidade de o tamanho do grão exceder 100?

b. Qual é a probabilidade de o tamanho do grão estar entre 50 e 80?

c. Que intervalo (a, b) inclui os 90% centrais de todos os tamanhos de grãos (de forma que 5% estejam abaixo de a e 5% acima de b)?

100. O tempo de reação (em segundos) a um determinado estímulo é uma variável aleatória contínua com fdp

$$f(x) = \begin{cases} \frac{3}{2} \cdot \frac{1}{x^2} & 1 \leq x \leq 3 \\ 0 & \text{caso contrário} \end{cases}$$

a. Deduza a fdc.

b. Qual é a probabilidade de o tempo de reação ser no máximo 2,5 segundos? Entre 1,5 e 2,5 segundos?

c. Calcule o tempo de reação esperado.

d. Calcule o desvio padrão do tempo de reação.

e. Se um indivíduo levar mais de 1,5 segundo para reagir, uma luz se acende e fica acesa até que mais um segundo passe ou até a pessoa reagir (o que ocorrer primeiro). Determine o tempo esperado de luz acesa. [*Sugestão*: Seja $h(X)$ = tempo em que a luz fica acesa como função do tempo de reação X.]

101. Seja X a temperatura em que uma reação química acontece. Suponha que a fdp de X seja

$$f(x) = \begin{cases} \frac{1}{9}(4 - x^2) & -1 \leq x \leq 2 \\ 0 & \text{caso contrário} \end{cases}$$

a. Faça o gráfico de $f(x)$.

b. Determine a fdc e faça o gráfico.

c. 0 é a temperatura mediana em que ocorre a reação? Caso não seja, essa temperatura mediana é maior ou menor que 0?

d. Suponha que a reação ocorra de forma independente, uma vez em cada um de 10 diferentes laboratórios e que a fdp do tempo de reação em cada laboratório seja fornecida. Seja Y = número entre os 10 laboratórios em que a temperatura excede 1. Qual é o tipo de distribuição de Y? (Forneça o nome e os valores dos parâmetros.)

102. O artigo "Determination of the MTF of Positive Photoresists Using the Monte Carlo Method" (*Photographic Sci. and Engr.*, 1983, p. 254-260) propõe uma distribuição exponencial com parâmetro $\lambda = 0{,}93$ como modelo de distribuição do comprimento da trajetória de um fóton (μm) sob determinadas circunstâncias. Suponha que esse seja o modelo correto.

a. Qual é o comprimento esperado da trajetória e qual é o desvio padrão?

b. Qual é a probabilidade de o comprimento da trajetória exceder 3,0? Qual é a probabilidade de o comprimento da trajetória estar entre 1,0 e 3,0?

c. Que valor é excedido por apenas 10% de todos os comprimentos de trajetórias?

103. O artigo "The Prediction of Corrosion by Statistical Analysis of Corrosion Profiles" (*Corrosion Science*, 1985, p. 305-315) sugere a fdc a seguir para a profundidade X da cavidade mais profunda do experimento que envolve a exposição de aço carbono manganês à água do mar acidificada.

$$F(x; \alpha, \beta) = e^{-e^{-(x-\alpha)/\beta}} \qquad -\infty < x < \infty$$

Os autores propõem os valores $\alpha = 150$ e $\beta = 90$. Assuma esse modelo como o correto.

a. Qual é a probabilidade de a profundidade da cavidade mais profunda ser no máximo 150? No máximo 300? Entre 150 e 300?

b. Abaixo de que valor a profundidade máxima do poço pode ser observada em 90% de todos os experimentos?

c. Qual é a função densidade de X?

d. A função densidade pode ser mostrada como sendo unimodal (um único pico). Acima de que valor do eixo das medidas ocorre o pico? (Esse valor é a moda.)

e. Podemos demonstrar que $E(X) \approx 0{,}5772\beta + \alpha$. Qual é a média dos valores fornecidos de α e β, e como ele se compara à mediana e à moda? Desenhe o gráfico da função densidade. (Nota: Ela é denominada distribuição de maior valor extremo.)

104. Um componente tem uma vida útil X com distribuição exponencial de parâmetro λ.

a. Se o custo operacional por unidade de tempo for c, qual será o custo operacional esperado desse componente durante sua vida útil?

b. Em vez de uma taxa de custo constante c como na parte (a), suponha que a taxa de custo seja $c(1 - 0{,}5e^{ax})$ com $a < 0$, de forma que o custo por unidade de tempo é menor do que c quando o componente é novo e fica mais caro à medida que o componente envelhece. Calcule agora o custo operacional esperado durante a vida útil do componente.

105. A moda de uma distribuição contínua é o valor x^* que maximiza $f(x)$.

a. Qual é a moda de uma distribuição normal com parâmetros μ e σ?

b. A distribuição uniforme com parâmetros A e B tem uma única moda? Por quê?

c. Qual é a moda de uma distribuição exponencial com parâmetro λ? (Desenhe uma ilustração.)

d. Se X possui uma distribuição gama com parâmetros β, e $\alpha > 1$, determine a moda. [*Sugestão*: $\ln[f(x)]$ será maximizada se, e somente se, $f(x)$ o for, e é mais simples obter a derivada de $\ln[f(x)]$.]

e. Qual é a moda de uma distribuição qui quadrado com ν graus de liberdade?

106. O artigo "Error Distribution in Navigation" (*J. Institute of Navigation*, 1971, p. 429-442) sugere que a distribuição de freqüência de erros positivos (magnitudes dos erros) é bem aproximada por uma distribuição exponencial. Seja X = o erro de posição lateral (milhas náuticas), que pode ser negativo ou positivo. Suponha que a fdp de X seja

$$f(x) = (0{,}1)e^{-0{,}2|x|} \quad -\infty < x < \infty$$

a. Desenhe um gráfico de $f(x)$ e demonstre que $f(x)$ é uma fdp legítima (mostre que a integral é 1).

b. Deduza a fdc de X e faça o gráfico.

c. Calcule $P(X \le 0)$, $P(X \le 2)$, $P(-1 \le X \le 2)$ e a probabilidade de se cometer um erro de mais de 2 milhas.

107. Em alguns sistemas, um cliente é alocado a uma de duas instalações. Se o tempo de serviço de um cliente servido por uma instalação i tiver uma distribuição com parâmetros λ_i $(i = 1, 2)$ e p for a proporção de todos os clientes servidos pela instalação 1, a fdp de X = tempo de serviço de um cliente selecionado aleatoriamente será

$$f(x; \lambda_1, \lambda_2, p) = \begin{cases} p\lambda_1 e^{-\lambda_1 x} + (1-p)\lambda_2 e^{-\lambda_2 x} & x \ge 0 \\ 0 & \text{caso contrário} \end{cases}$$

Essa distribuição é freqüentemente denominada distribuição exponencial mista ou hiperexponencial. Ela também é proposta como modelo para quantidade de chuva em "Modeling Monsoon Affected Rainfall of Pakistan by Point Processes" (*J. Water Resources Planning and Mgmnt.*, 1992, p. 671-688).

a. Demonstre que $f(x; \lambda_1, \lambda_2, p)$ realmente é uma fdp.

b. Qual é a fdc $F(x; \lambda_1, \lambda_2, p)$?

c. Se X tiver $f(x; \lambda_1, \lambda_2, p)$ como sua fdp, qual será $E(X)$?

d. Usando o fato de que $E(X^2) = 2/\lambda^2$ quando X possui uma distribuição exponencial com parâmetro λ, calcule $E(X^2)$ quando X tiver fdp $f(x; \lambda_1, \lambda_2, p)$. Depois, calcule $V(X)$.

e. O coeficiente da variação de uma variável aleatória (ou distribuição) é $CV = \sigma/\mu$. Qual é o CV para uma va exponencial? O que você pode dizer sobre o valor CV quando X tem uma distribuição hiperexponencial?

f. Qual é o CV de uma distribuição de Erlang com parâmetros λ e n como definido no Exercício 62? (*Nota*: No trabalho aplicado, o CV amostral é usado para decidir qual das três distribuições pode ser apropriada.)

108. Suponha que um estado específico permita que os indivíduos preencham restituições de imposto somente se o total das deduções em itens ultrapassar valores acima de US$ 5000. Seja X (em milhares de dólares) o total das deduções em itens de um formulário escolhido aleatoriamente. Assuma que X tenha fdp

$$f(x; \alpha) = \begin{cases} k/x^{\alpha} & x \geq 5 \\ 0 & \text{caso contrário} \end{cases}$$

a. Determine o valor de k. Que restrição é necessária para α?

b. Qual é a fdc de X?

c. Qual é a dedução total esperada de um formulário escolhido aleatoriamente? Que restrição é necessária para α tal que $E(X)$ seja finito?

d. Mostre que $\ln(X/5)$ possui uma distribuição exponencial com parâmetro $\alpha - 1$.

109. Seja por I_i a corrente de entrada de um transistor e I_0 a corrente de saída. O ganho de corrente então será proporcional a $\ln(I_0/I_i)$. Suponha que a constante de proporcionalidade seja 1 (que implica a escolha de uma unidade de medida específica), de forma que o ganho da corrente $= X = \ln(I_0/I_i)$. Assuma X como distribuição normal com $\mu = 1$ e $\sigma = 0{,}05$.

a. Que tipo de distribuição tem a razão I_0/I_i?

b. Qual é a probabilidade de a corrente de saída ser maior que o dobro da corrente de entrada?

c. Qual é o valor esperado e a variância da razão da corrente de saída partir a de entrada?

110. O artigo "Response of SiC_f/Si_3N_4, Composites Under Static and Cyclic Loading — An Experimental and Statistical Analysis" (*J. of Engr. Materials and Technology*, 1997, p. 186-193) sugere que a resistência à tração (MPa) de compostos em condições específicas pode ser modelada por uma distribuição de Weibull com $\alpha = 9$ e $\beta = 180$.

a. Desenhe o gráfico da função de densidade.

b. Qual é a probabilidade de a resistência de um espécime selecionado aleatoriamente exceder 175? Estar entre 150 e 175?

c. Se dois espécimes selecionados aleatoriamente forem escolhidos de forma independente, qual será a probabilidade de ao menos um ter uma resistência entre 150 e 175?

d. Que valor de resistência separa os 10% mais fracos de todos os espécimes dos 90% restantes?

111. Assuma Z com uma distribuição normal padrão e defina uma nova va Y como $Y = \sigma Z + \mu$. Mostre que Y tem distribuição normal com parâmetros μ e σ. (*Sugestão*: $Y \leq y$ se, e somente se, $Z \leq$? Use isso para determinar a fdc de Y e depois diferencie em relação a y.)

112. a. Suponha que a vida útil X de um componente, quando medido em horas, tenha uma distribuição gama com parâmetros α e β. Seja $Y =$ vida útil medida em minutos. Deduza a fdp de Y. (*Sugestão*: $Y \leq y$ se e somente $X \leq y/60$. Use isso para obter a fdc de Y e depois diferencie para obter a fdp.)

b. Se X tem distribuição gama com parâmetros α e β, qual é a distribuição probabilidade $Y = cX$?

113. Nos Exercícios 111 e 112, bem como em muitas outras situações, temos a fdp $f(x)$ de X e queremos saber a fdp de $Y = h(X)$. Assuma que $h(\cdot)$ é uma função inversível, de forma que $y = h(x)$ pode ser resolvida para x para se obter $x = k(y)$. Podemos então mostrar que a fdp de Y é

$$g(y) = f[k(y)] \cdot |k'(y)|$$

a. Se X tiver distribuição uniforme com $A = 0$ e $B = 1$, encontre a fdp de $Y = -\ln(X)$.

b. Resolva o Exercício 111 usando esse resultado.

c. Resolva o Exercício 112 (b) usando esse resultado.

114. Com base nos dados de um experimento de lançamento de dardos, o artigo "Shooting Darts" (*Chance*, Summer, 1997, p. 16-19) propôs que os erros horizontais e verticais em relação a um alvo pontual devem ser independentes um do outro, cada um com distribuição normal de média 0 e variância σ^2. Podemos mostrar então que a fdp da distância V do alvo até o ponto de toque é

$$f(v) = \frac{v}{\sigma^2} \cdot e^{-v^2/2\sigma^2} \quad v > 0$$

a. Essa fdp é membro de que família apresentada neste capítulo?

b. Se $\sigma = 20$mm (próximo ao valor sugerido no artigo), qual é a probabilidade de um dardo ser acertado dentro de 25mm (cerca de 1 pol.) do alvo?

115. O artigo "Three Sisters Give Birth on the Same Day" (*Chance*, Spring 2001, p. 23-25) usou o fato de três irmãs de Utah darem à luz em 11 de março de 1998 como base de proposta de algumas questões interessantes relativas a coincidências de nascimento.

a. Desconsiderando anos bissextos e assumindo que os outros 365 dias são igualmente prováveis, qual é a probabilidade de que três nascimentos selecionados aleatoriamente ocorram todos em 11 de março? Indique quaisquer hipóteses existentes, se houver.

b. Com as hipóteses usadas na parte (a), qual é a probabilidade de três nascimentos selecionados aleatoriamente ocorrerem no mesmo dia?

c. O autor sugeriu que, com base em extensos dados, a duração da gestação (tempo entre a concepção e o nascimento) pode ser modelada como tendo

distribuição normal com valor médio 280 dias e desvio padrão 19,88 dias. As datas-limite para as três irmãs de Utah eram 15 de março, 1° de abril e 4 de abril, respectivamente. Supondo que o prazo esteja na média da distribuição, qual é a probabilidade de todos os nascimentos ocorrerem em 11 de março? (*Sugestão*: O desvio em relação à data de nascimento tem distribuição normal com média 0.)

d. Explique como você usaria as informações da parte (c) para calcular a probabilidade de uma data de nascimento comum.

116. Seja X a vida útil de um componente, com $f(x)$ e $F(x)$ sendo a fdp e a fdc de X. A probabilidade de o componente apresentar falha no intervalo $(x, x + \Delta x)$ é aproximadamente $f(x) \cdot \Delta x$. A probabilidade condicional de o componente apresentar falha em $(x, x + \Delta x)$, dado que durou ao menos x, é $f(x) \cdot \Delta x/[1 - F(x)]$. Dividindo a expressão por Δx produz a **função de taxa de falha**:

$$r(x) = \frac{f(x)}{1 - F(x)}$$

Um aumento da função de taxa de falha indica que os componentes mais antigos têm mais probabilidade de apresentar desgaste, enquanto uma taxa de falha decrescente é evidência de aumento de confiabilidade com a idade. Na prática, uma falha "com formato de banheira" é assumida com freqüência.

a. Se X tiver distribuição exponencial, qual será $r(x)$?

b. Se X tiver distribuição de Weibull com parâmetros α e β, qual será $r(x)$? Para que valores de parâmetros $r(x)$ aumentará? Para que valores de parâmetros $r(x)$ diminuirá com x?

c. Como $r(x) = -(d/dx)\ln[1 - F(x)]$, $\ln[1 - F(x)] = -\int r(x)\,dx$. Suponha

$$r(x) = \begin{cases} \alpha\left(1 - \dfrac{x}{\beta}\right) & 0 \le x \le \beta \\ 0 & \text{caso contrário} \end{cases}$$

de forma que, se um componente durar β horas, durará para sempre (apesar de não parecer razoável, o modelo é usado para estudar apenas o "desgaste inicial"). Quais são a fdc e a fdp de X?

117. Assuma U como tendo uma distribuição uniforme no intervalo [0, 1]. Os valores observados com essa distribuição podem ser obtidos com um gerador de números aleatórios. Seja $X = -(1/\lambda)\ln(1 - U)$.

a. Mostre que X possui uma distribuição exponencial com parâmetro λ. [Sugestão: a fdc de X é $F(x) = P(X \le x)$; $X \le x$ é equivalente a $U \le$?]

b. Como você usaria a parte (a) e um gerador de números aleatórios para obter valores observados de uma distribuição exponencial com parâmetro $\lambda = 10$?

118. Considere uma va X com média μ e desvio padrão σ e seja $g(X)$ uma função especificada de X. A série de aproximação de Taylor de primeira ordem em relação a $g(X)$ na vizinhança de μ é

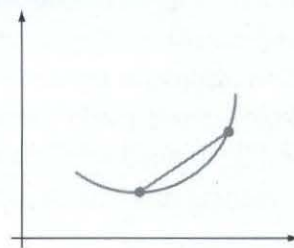
$$g(X) \approx g(\mu) + g'(\mu) \cdot (X - \mu)$$

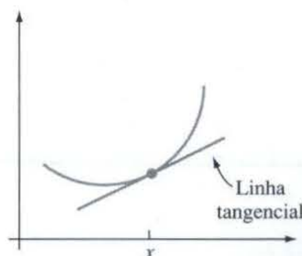
O lado direito dessa equação é uma função linear de X. Se a distribuição de X estiver concentrada em um intervalo em que $g(\cdot)$ é aproximadamente linear [por exemplo, $\sqrt{x}$ é aproximadamente linear em (1, 2)], a equação fornece aproximações para $E(g(X))$ e $V(g(X))$.

a. Forneça expressões para tais aproximações. (*Sugestão*: use regras de valor esperado e variância para uma função linear $aX + b$.).

b. Se a voltagem v em um meio for fixa mas a corrente I for aleatória, então a resistência também será uma variável aleatória relacionada a I por $R = v/I$. Se $\mu_I = 20$ e $\sigma_I = 0{,}5$, calcule aproximações para μ_R e σ_R.

119. Uma função $g(x)$ é convexa se a corda que conecta quaisquer dois pontos no gráfico da função ficar acima do gráfico. Quando $g(x)$ é diferenciável, uma condição equivalente é que, para cada x, a reta tangente em x esteja totalmente acima ou abaixo do gráfico. (Veja as figuras abaixo.) Como $g(\mu) = g(E(X))$ se compara a $E(g(X))$? [*Sugestão*: a equação da tangente em $x = \mu$ é $y = g(\mu) + g'(\mu) \cdot (x - \mu)$. Use a condição de convexidade, substitua X por x e calcule os valores esperados. *Nota*: a menos que $g(x)$ seja linear, a desigualdade resultante (normalmente denominada desigualdade de Jensen) é estrita ($<$ em vez de $\le$). Isso é válido para vas contínuas e discretas.]





120. Assuma X com distribuição de Weibull de parâmetros $\alpha = 2$ e β. Mostre que $Y = 2X^2/\beta^2$ possui uma distribuição qui quadrado com $\nu = 2$. [*Sugestão*: a fdc de Y é $P(Y \le y)$. Expresse essa probabilidade sob a forma $P(X \le g(y))$, usando o fato de que X possui fdc da forma da Expressão (4.12) e diferencie em relação à y para obter a fdp de Y.]

Bibliografia

BURY, Karl, *Statistical Distributions in Engineering*, Cambridge: Cambridge Univ. Press, 1999. Uma pesquisa informativa e de fácil leitura sobre distribuições e suas propriedades.

JOHNSON, Norman; KOTZ, Samuel e BALAKRISHNAN N. *Continuous Univariate Distributions*, Wiley: Nova York, 1994, vols. 1–2. Esses dois volumes apresentam uma pesquisa abrangente sobre as diversas distribuições contínuas.

NELSON, Wayne. *Applied Life Data Analysis*, Wiley. Nova York, 1982. Fornece uma discussão abrangente das distribuições e dos métodos usados na análise de dados de vida útil.

OLKIN, Ingram, DERMAN, Cyrus e GLESER, Leon, *Probability Models and Applications*. Macmillan: Nova York, 1994, (2 ed.). Boa cobertura de propriedades gerais e de distribuições específicas.

5

Distribuições de Probabilidade Conjunta e Amostras Aleatórias

Introdução

Nos capítulos 3 e 4, estudamos modelos probabilísticos com uma única variável aleatória. Muitos problemas de probabilidade e estatística nos conduzem a modelos que envolvem diversas variáveis aleatórias simultaneamente. Neste capítulo, discutiremos inicialmente os modelos probabilísticos para o comportamento conjunto de diversas variáveis aleatórias, enfatizando o caso em que as variáveis são independentes umas das outras. Depois estudaremos os valores esperados das funções de diversas variáveis aleatórias, incluindo covariância e correlação como medida do grau de relação entre duas variáveis.

As três últimas seções do capítulo consideram funções de n variáveis aleatórias $X_1, X_2, \ldots, X_n$, enfocando especialmente as médias $(X_1 + \cdots + X_n)/n$. Denominamos as funções desse tipo, que são elas mesmas variáveis aleatórias, uma estatística. Resultados da probabilidade são usados para obtenção de informações sobre a distribuição de uma estatística. O resultado principal desse tipo é o Teorema do Limite Central (CLT), a base de vários procedimentos inferenciais que envolvem tamanhos de amostras grandes.

5.1 Variáveis Aleatórias de Distribuição Conjunta

Há várias situações experimentais em que mais de uma variável aleatória será de interesse do investigador. Primeiro consideraremos as distribuições de probabilidade conjunta para duas vas discretas, depois para duas vas contínuas e finalmente para mais de duas variáveis.

A Função de Massa de Probabilidade Conjunta Para Duas Variáveis Aleatórias Discretas

A função de massa de probabilidade (fmp) de uma única va discreta X especifica quanta massa de probabilidade é colocada em cada valor X possível. A fmp conjunta de duas vas discretas X e Y descreve quanta massa de probabilidade é colocada em cada par de valores possível (x, y).

DEFINIÇÃO

Sejam X e Y duas vas discretas definidas no espaço amostral $\mathcal{S}$ de um experimento. A **função de massa de probabilidade conjunta** $p(x, y)$ é definida para cada par de números (x, y) por

$$p(x, y) = P(X = x \text{ e } Y = y)$$

Seja A qualquer conjunto formado por pares de valores (x, y). A probabilidade $P[(X, Y) \in A]$ é obtida pela soma da fmp conjunta com os pares de A:

$$P[(X, Y) \in A] = \sum_{(x, y) \in A} \sum p(x, y)$$

Exemplo 5.1

Uma grande agência de seguros presta serviços a diversos clientes que compraram uma apólice residencial e outra de automóvel da mesma seguradora. Para cada tipo, deve ser especificado um valor dedutível. Para uma apólice de automóvel as opções são US\$ 100 e US\$ 250, enquanto, para uma apólice residencial, as opções são 0, US\$ 100 e US\$ 200. Suponha que um indivíduo com os dois referidos tipos seja selecionado aleatoriamente nos arquivos da seguradora. Sejam X = valor dedutível na apólice de automóvel e Y = valor dedutível na apólice residencial. Os pares (X, Y) possíveis são (100, 0), (100, 100), (100, 200), (250, 0), (250, 100) e (250, 200); a fmp conjunta especifica a probabilidade associada a cada um desses pares, com qualquer outro par de probabilidade zero. Suponha que a fmp conjunta seja dada na **tabela de probabilidade conjunta** a seguir:

$p(x, y)$		y: 0	100	200
x	100	0,20	0,10	0,20
	250	0,05	0,15	0,30

Então, $p(100, 100) = P(X = 100 \text{ e } Y = 100) = P(\text{US\$100 dedutível nas duas apólices}) = 0{,}10$. A probabilidade $P(Y \geq 100)$ é calculada pela soma das probabilidades de todos os pares (x, y) para os quais $y \geq 100$:

$$\begin{aligned} P(Y \geq 100) &= p(100, 100) + p(250, 100) + p(100, 200) + p(250, 200) \\ &= 0{,}75 \end{aligned}$$

■

Uma função $p(x, y)$ pode ser usada como fmp conjunta desde que $p(x, y) \geq 0$ para todos os x e y e $\sum_x \sum_y p(x, y) = 1$.

A fmp de apenas uma das variáveis é obtida pela soma de $p(x, y)$ em relação aos valores da outra variável. O resultado é denominado *fmp marginal* porque, quando os valores $p(x, y)$ são exibidos em uma tabela retangular, as somas são apenas totais marginais (linha ou coluna).

DEFINIÇÃO

As **funções de massa de probabilidade marginais** de X e de Y, representadas respectivamente por $p_X(x)$ e $p_Y(y)$, são dadas por

$$p_X(x) = \sum_y p(x, y) \qquad p_Y(y) = \sum_x p(x, y)$$

Assim, para obter a fmp marginal de X calculado, digamos, para $x = 100$, as probabilidades $p(100, y)$ são somadas para todos os valores possíveis y. Cálculo de cada X possível fornece a fmp marginal de X apenas (sem referência a Y). Pelas fmps marginais, podem ser calculadas as probabilidades de eventos que envolvem apenas X ou apenas Y.

Exemplo 5.2 (continuação do Exemplo 5.1)

Os valores de X possíveis são $x = 100$ e $x = 250$, assim o cálculo dos totais de linhas da tabela de probabilidade conjunta fornece

$$p_X(100) = p(100, 0) + p(100, 100) + p(100, 200) = 0{,}50$$

e

$$p_X(250) = p(250, 0) + p(250, 100) + p(250, 200) = 0{,}50$$

A fmp marginal de X será então

$$p_X(x) = \begin{cases} 0{,}5 & x = 100, 250 \\ 0 & \text{caso contrário} \end{cases}$$

De forma similar, a fmp marginal de Y será obtida pelos totais da coluna como a seguir

$$p_Y(y) = \begin{cases} 0{,}25 & y = 0, 100 \\ 0{,}50 & y = 200 \\ 0 & \text{caso contrário} \end{cases}$$

então $P(Y \geq 100) = p_Y(100) + p_Y(200) = 0{,}75$ como antes. ■

A Função de Densidade de Probabilidade Conjunta para Duas Variáveis Aleatórias Contínuas

A probabilidade de o valor observado de uma va contínua X estar em um conjunto unidimensional A (como um intervalo) é obtida integrando-se a fdp $f(x)$ em relação ao conjunto A. De forma similar, a probabilidade de o par (X, Y) de vas contínuas estar em um conjunto bidimensional A (como um retângulo) é obtida pela integração de uma função denominada *função de densidade conjunta.*

DEFINIÇÃO

Seja X e Y vas contínuas. Então, $f(x, y)$ é a **função de densidade de probabilidade conjunta** de X e Y se, para qualquer conjunto bidimensional, A

$$P[(X, Y) \in A] = \iint_A f(x, y)\, dx\, dy$$

Em particular, se A for o retângulo bidimensional $\{(x, y): a \leq x \leq b, c \leq y \leq d\}$, então

$$P[(X, Y) \in A] = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\, dy\, dx$$

Para $f(x, y)$ ser candidata a uma fdp conjunta, deve satisfazer $f(x, y) \geq 0$ e $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$. Podemos imaginar $f(x, y)$ como especificação de uma superfície de altura $f(x, y)$ acima do ponto (x, y) em um sistema tridimensional. Então $P[(X, Y) \in A]$ é o volume abaixo dessa superfície e acima da região A, análogo à área abaixo de uma curva no caso unidimensional. Como ilustrado na Figura 5.1.

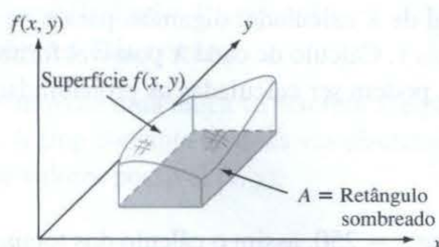


Figura 5.1 $P[(X, Y) \in A]$ = volume sob a superfície de densidade acima de A

Exemplo 5.3

Um banco opera tanto uma instalação de *drive-through* como em guichê de atendimento. Em um dia selecionado aleatoriamente, assuma X = a proporção de tempo em que a instalação de *drive-through* está em uso (ao menos um cliente está sendo atendido ou esperando para ser atendido) e Y = a proporção de tempo em que o guichê de atendimento está em uso. O conjunto de valores possíveis de (X, Y) é, então, o retângulo $D = \{(x, y): 0 \le x \le 1, 0 \le y \le 1\}$. Suponha que a fdp conjunta de (X, Y) seja dada por

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{caso contrário} \end{cases}$$

Para provar que essa fdp é verdadeira, observe que $f(x, y) \ge 0$ e

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy &= \int_0^1\int_0^1 \frac{6}{5}(x + y^2)\,dx\,dy \\ &= \int_0^1\int_0^1 \frac{6}{5}x\,dx\,dy + \int_0^1\int_0^1 \frac{6}{5}y^2\,dx\,dy \\ &= \int_0^1 \frac{6}{5}x\,dx + \int_0^1 \frac{6}{5}y^2\,dy = \frac{6}{10} + \frac{6}{15} = 1 \end{aligned}$$

A probabilidade de nenhuma das instalações estar ocupada em mais de um quarto do tempo é

$$\begin{aligned} P\left(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}\right) &= \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x + y^2)\,dx\,dy \\ &= \frac{6}{5}\int_0^{1/4}\int_0^{1/4} x\,dx\,dy + \frac{6}{5}\int_0^{1/4}\int_0^{1/4} y^2\,dx\,dy \\ &= \frac{6}{20}\cdot\frac{x^2}{2}\bigg|_{x=0}^{x=1/4} + \frac{6}{20}\cdot\frac{y^3}{3}\bigg|_{y=0}^{y=1/4} = \frac{7}{640} \\ &= 0{,}0109 \end{aligned}$$

■

Como acontece com as fmps conjuntas, cada uma das duas funções densidade marginais pode ser calculada pela fdp de X e Y.

DEFINIÇÃO As **funções de densidade de probabilidade marginais** de X e Y, representadas respectivamente por $f_X(x)$ e $f_Y(y)$, são dadas por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy \qquad \text{para } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx \qquad \text{para } -\infty < y < \infty$$

Exemplo 5.4 (continuação do Exemplo 5.3)

A fdp marginal de X, que fornece a distribuição de probabilidades do tempo ocupado para a instalação de *drive-through* sem referência à janela de atendimento, é

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^1 \frac{6}{5}(x + y^2)\, dy = \frac{6}{5}x + \frac{2}{5}$$

para $0 \leq x \leq 1$ e 0 em caso contrário. A fdp marginal de Y é

$$f_Y(y) = \begin{cases} \frac{6}{5}y^2 + \frac{3}{5} & 0 \leq y \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Então

$$P\left(\frac{1}{4} \leq Y \leq \frac{3}{4}\right) = \int_{1/4}^{3/4} f_Y(y)\, dy = \frac{37}{80} = 0{,}4625$$ ■

No Exemplo 5.3, a região da densidade conjunta positiva era um retângulo, o que tornou o cálculo das fdps marginais relativamente fácil. Considere agora um exemplo em que a região de densidade positiva é uma figura mais complicada.

Exemplo 5.5

Uma empresa de nozes comercializa latas luxuosas de nozes mistas com amêndoas, castanhas de caju e amendoins. Suponha que o peso líquido de cada lata seja exatamente 1 libra, mas que a contribuição do peso de cada tipo de noz seja aleatória. Como os três pesos devem somar 1, um modelo de probabilidade conjunta para quaisquer dois fornece todas as informações necessárias sobre o peso do terceiro tipo. Sejam X = peso das amêndoas em uma lata selecionada e Y = peso das castanhas de caju. Então a região de densidade positiva é $D = \{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1\}$, região sombreada ilustrada na Figura 5.2.

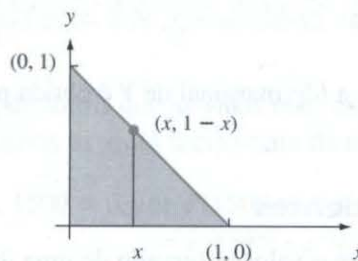


Figura 5.2 Região de densidade positiva do Exemplo 5.5

Agora considere que a fdp conjunta de (X, Y) seja

$$f(x, y) = \begin{cases} 24xy & 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Para qualquer x fixo, $f(x, y)$ aumenta com y; para y fixo, $f(x, y)$ aumenta com x. A hipótese é apropriada porque a palavra *luxo* implica que a maior parte da lata deve ser formada por amêndoas e castanhas de caju em vez de amendoins, de forma que a função de densidade deve ser grande próximo ao limite superior e pequena perto da origem. A superfície determinada por $f(x, y)$ tem declive positivo a partir de zero, conforme (x, y) se distancia dos eixos.

Claramente, $f(x, y) \geq 0$. Para demonstrar a segunda condição de uma fdp conjunta, lembre-se de que a integral dupla é calculada como integral iterativa, mantendo uma variável fixa (como x na Figura 5.2),

integrando em relação aos valores da outra variável que estão na reta que passa pelo valor da variável fixa e, finalmente, integrando em relação a todos os valores possíveis da variável fixa. Dessa forma,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dy\, dx = \int_D\int f(x, y)\, dy\, dx = \int_0^1\left\{\int_0^{1-x} 24xy\, dy\right\} dx$$

$$= \int_0^1 24x\left\{\frac{y^2}{2}\Bigg|_{y=0}^{y=1-x}\right\} dx = \int_0^1 12x(1-x)^2\, dx = 1$$

Para calcular a probabilidade de dois tipos de nozes formarem juntos no máximo 50% da lata, assuma $A = \{(x, y): 0 \le x \le 1, 0 \le y \le 1, \text{e } x + y \le 0{,}5\}$, conforme mostrado na Figura 5.3. Então

$$P((X, Y) \in A) = \int_A\int f(x, y)\, dx\, dy = \int_0^{0,5}\int_0^{0,5-x} 24xy\, dy\, dx = 0{,}0625$$

A fdp marginal das amêndoas é obtida fixando-se X em x e integrando $f(x, y)$ ao longo da linha vertical que passa por x:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \begin{cases} \int_0^{1-x} 24xy\, dy = 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & \text{caso contrário} \end{cases}$$

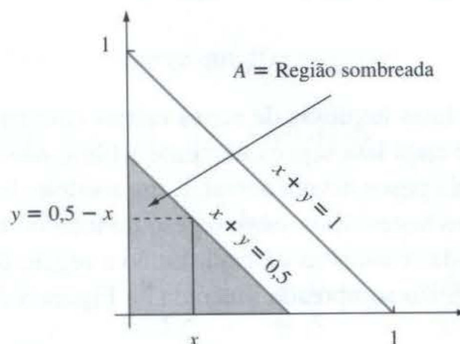


Figura 5.3 Cálculo de $P[(X, Y) \in A]$ do Exemplo 5.5

Pela simetria de $f(x, y)$ e a região D, a fdp marginal de Y é obtida pela substituição de x e X em $f_X(x)$ por y e Y, respectivamente.

Variáveis Aleatórias Independentes

Em muitas situações, as informações sobre o valor observado de uma das duas variáveis X e Y fornecem dados sobre o valor da outra variável. No Exemplo 5.1, a probabilidade marginal de X em $x = 250$ era 0,5, como a probabilidade de $X = 100$. Se, entretanto, dissessem-nos que a pessoa selecionada tinha $Y = 0$, então, $X = 100$ é quatro vezes mais provável que $X = 250$. Portanto, há uma dependência entre as duas variáveis.

No Capítulo 2, apontamos que uma forma de definir a independência de dois eventos é dizer que A e B são independentes, se $P(A \cap B) = P(A) \cdot P(B)$. Eis uma definição análoga da independência de duas vas.

DEFINIÇÃO

Duas variáveis aleatórias X e Y são **independentes** se, para cada par de valores x e y,

$$p(x, y) = p_X(x) \cdot p_Y(y) \quad \text{quando } X \text{ e } Y \text{ são discretas} \tag{5.1}$$

ou

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \text{quando } X \text{ e } Y \text{ são contínuas}$$

Se (5.1) não for satisfeita para todos os (x, y), então, X e Y são denominadas **dependentes**.

A definição diz que duas variáveis são independentes se sua fmp conjunta ou sua fdp for o produto das duas fmps ou fdps marginais.

Exemplo 5.6

No caso do seguro dos Exemplos 5.1 e 5.2,

$$p(100, 100) = 0{,}10 \neq (0{,}5)(0{,}25) = p_X(100) \cdot p_Y(100)$$

de forma que X e Y não são independentes. A independência de X e Y exige que *cada* entrada na tabela de probabilidade conjunta seja o produto das probabilidades marginais da linha e da coluna correspondentes. ■

Exemplo 5.7 (continuação do Exemplo 5.5)

Como $f(x, y)$ tem a forma de um produto, X e Y podem parecer independentes. Entretanto, apesar de $f_X(\frac{3}{4}) = f_Y(\frac{3}{4}) = \frac{9}{16}$, $f(\frac{3}{4}, \frac{3}{4}) = 0 \neq \frac{9}{16} \cdot \frac{9}{16}$, então as variáveis não são de fato independentes. Para serem independentes, $f(x, y)$ deve ter a forma $g(x) \cdot h(y)$ *e* a região de densidade positiva deve ser um retângulo cujos lados são paralelos aos eixos de coordenadas. ■

A independência de duas variáveis aleatórias é mais útil quando a descrição do experimento em estudo nos diz que X e Y não afetam uma à outra. Então, uma vez especificadas as fdps ou as fmps marginais, a fmp ou a fdp conjunta é simplesmente o produto das duas funções marginais. Segue que

$$P(a \leq X \leq b, c \leq Y \leq d) = P(a \leq X \leq b) \cdot P(c \leq Y \leq d)$$

Exemplo 5.8

Suponha que as vidas úteis de dois componentes sejam independentes uma da outra e que a primeira vida útil, X_1, tenha distribuição exponencial com parâmetro λ_1, enquanto a segunda, X_2, tenha distribuição exponencial com parâmetro λ_2. A fdp conjunta será então

$$f(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

$$= \begin{cases} \lambda_1 e^{-\lambda_1 x_1} \cdot \lambda_2 e^{-\lambda_2 x_2} = \lambda_1 \lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2} & x_1 > 0, x_2 > 0 \\ 0 & \text{caso contrário} \end{cases}$$

Sejam $\lambda_1 = 1/1000$ e $\lambda_2 = 1/1200$, de forma que as vidas úteis esperadas sejam de 1000 e 1200 horas, respectivamente. A probabilidade de que ambas as vidas úteis sejam de no mínimo 1500 horas é

$$\begin{aligned} P(1500 \leq X_1, 1500 \leq X_2) &= P(1500 \leq X_1) \cdot P(1500 \leq X_2) \\ &= e^{-\lambda_1(1500)} \cdot e^{-\lambda_2(1500)} \\ &= (0{,}2231)(0{,}2865) = 0{,}0639 \end{aligned}$$

■

Mais de Duas Variáveis Aleatórias

Para modelar o comportamento conjunto de mais de duas variáveis aleatórias, estendemos o conceito de distribuição conjunta de duas variáveis.

DEFINIÇÃO

Se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias discretas, a fmp conjunta das variáveis é a função

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Se as variáveis forem contínuas, a fdp conjunta de $X_1, \ldots, X_n$ será a função $f(x_1, x_2, \ldots, x_n)$ de forma que, para qualquer dos n intervalos $[a_1, b_1], \ldots, [a_n, b_n]$,

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n) = \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \ldots dx_1$$

Em um experimento binomial, cada tentativa pode resultar em um de somente dois resultados possíveis. Considere agora um experimento formado por n tentativas idênticas e independentes, em que cada tentativa pode resultar em um dos r resultados possíveis. Seja $p_i = P(\text{resultado } i \text{ em qualquer tentativa específica})$ e defina as variáveis aleatórias por X_i = o número de tentativas que resultam em i ($i = 1, \ldots, r$). Tal experimento é denominado **experimento multinomial**, e a fmp conjunta de $X_1, \ldots, X_r$, é denominada **distribuição multinomial**. Usando um raciocínio de contagem análogo ao usado na demonstração da distribuição binomial, a fmp conjunta de $X_1, \ldots, X_r$ pode ser demonstrada

$$p(x_1, \ldots, x_r) = \begin{cases} \dfrac{n!}{(x_1!)(x_2!)\cdot\cdots\cdot(x_r!)} p_1^{x_1}\cdot\cdots\cdot p_r^{x_r} & x_i = 0, 1, 2, \ldots, \text{com } x_1 + \cdots + x_r = n \\ 0 & \text{caso contrário} \end{cases}$$

O caso $r = 2$ fornece a distribuição binomial, com X_1 = número de sucessos e $X_2 = n - X_1$ = número de falhas.

Exemplo 5.9

Se o alelo de cada uma das 10 seções da ervilha obtidas de forma independente é determinado e $p_1 = P(\text{AA})$, $p_2 = P(\text{Aa})$, $p_3 = P(\text{aa})$, X_1 = número de AAs, X_2 = número de Aas e X_3 = número de aas, então

$$p(x_1, x_2, x_3) = \frac{10!}{(x_1!)(x_2!)(x_3!)} p_1^{x_1} p_2^{x_2} p_3^{x_3} \qquad x_i = 0, 1, \ldots \text{ e } x_1 + x_2 + x_3 = 10$$

Se $p_1 = p_3 = 0{,}25$, $p_2 = 0{,}5$, então

$$P(X_1 = 2, X_2 = 5, X_3 = 3) = p(2, 5, 3)$$
$$= \frac{10!}{2!\,5!\,3!}(0{,}25)^2(0{,}5)^5(0{,}25)^3 = 0{,}0769$$ ■

Exemplo 5.10

Quando um determinado método é usado para coletar um volume fixo de amostras de rochas de uma região, há quatro tipos de rochas resultantes. Represente por X_1, X_2, e X_3 a proporção por volume dos tipos de rochas 1, 2 e 3 em uma amostra selecionada aleatoriamente (a proporção da rocha do tipo 4 é $1 - X_1 - X_2 - X_3$, de forma que uma variável X_4 seria redundante). Se a fdp conjunta de X_1, X_2, X_3 for

$$f(x_1, x_2, x_3) = \begin{cases} kx_1x_2(1 - x_3) & 0 \le x_1 \le 1, 0 \le x_2 \le 1, 0 \le x_3 \le 1, x_1 + x_2 + x_3 \le 1 \\ 0 & \text{caso contrário} \end{cases}$$

então k será determinado por

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2, x_3)\, dx_3\, dx_2\, dx_1$$
$$= \int_0^1 \left\{\int_0^{1-x_1} \left[\int_0^{1-x_1-x_2} kx_1x_2(1 - x_3)\, dx_3\right] dx_2\right\} dx_1$$

Essa integral iterativa tem valor $k/144$, portanto, $k = 144$. A probabilidade de que as rochas dos tipos 1 e 2 juntas sejam responsáveis por no máximo 50% da amostra é

$$P(X_1 + X_2 \le 0{,}5) = \iiint\limits_{\substack{0 \le x_i \le 1 \text{ para } i = 1,2,3 \\ x_1 + x_2 + x_3 \le 1,\ x_1 + x_2 \le 0{,}5}} f(x_1, x_2, x_3)\, dx_3\, dx_2\, dx_1$$
$$= \int_0^{0{,}5} \left\{\int_0^{0{,}5-x_1} \left[\int_0^{1-x_1-x_2} 144x_1x_2(1 - x_3)\, dx_3\right] dx_2\right\} dx_1$$
$$= 0{,}6066$$ ■

A noção de independência de mais de duas variáveis aleatórias é semelhante à noção de independência de mais de dois eventos.

DEFINIÇÃO

> As variáveis aleatórias $X_1, X_2, \ldots, X_n$ são **independentes** se para *cada* subconjunto $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ das variáveis (cada par, cada trio e assim por diante), a fmp ou fdp conjuntas do subconjunto for igual ao produto das fdps ou fmps marginais.

Assim, se as variáveis forem independentes com $n = 4$, a fmp ou fdp conjunta de quaisquer duas variáveis será o produto das duas marginais e de forma semelhante para quaisquer três e todas as quatro variáveis. O mais importante é, depois que soubermos que n variáveis são independentes, saber que a fmp ou fdp conjunta será o produto das n marginais.

Exemplo 5.11

Se $X_1, \ldots, X_n$ representam as vidas úteis de n componentes que operam de forma independente um do outro e cada vida útil é distribuída exponencialmente com parâmetro λ, então

$$f(x_1, x_2, \ldots, x_n) = (\lambda e^{-\lambda x_1}) \cdot (\lambda e^{-\lambda x_2}) \cdot \cdots \cdot (\lambda e^{-\lambda x_n})$$

$$= \begin{cases} \lambda^n e^{-\lambda \Sigma x_i} & x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

Se esses n componentes formam um sistema que apresentará falha assim que um único componente tiver uma falha, a probabilidade de o sistema durar mais do que o tempo t é

$$P(X_1 > t, \ldots, X_n > t) = \int_t^\infty \cdots \int_t^\infty f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

$$= \left(\int_t^\infty \lambda e^{-\lambda x_1}\, dx_1\right) \cdots \left(\int_t^\infty \lambda e^{-\lambda x_n}\, dx_n\right)$$

$$= (e^{-\lambda t})^n = e^{-n\lambda t}$$

Portanto,

$$P(\text{tempo de vida do sistema} \leq t) = 1 - e^{-n\lambda t} \qquad \text{para } t \geq 0$$

que mostra que a vida útil do *sistema* tem uma distribuição exponencial com parâmetro $n\lambda$; e o valor esperado do tempo de vida do sistema é $1/n\lambda$. ■

Em muitas situações experimentais a serem consideradas neste livro, a independência é uma suposição razoável, de forma que especificar a distribuição conjunta se reduz a decidir quais são as distribuições marginais apropriadas.

Distribuições Condicionais

Suponha que X = número de defeitos graves em um novo automóvel selecionado aleatoriamente e Y = número de defeitos menores no mesmo carro. Sabendo que o número de defeitos graves no automóvel selecionado é um, qual é a probabilidade de o carro ter no máximo três defeitos menores, isto é, qual é $P(Y \leq 3 \mid X = 1)$? De forma semelhante, se X e Y representam as vidas úteis de dois componentes em um sistema e acontece de $X = 100$, qual é a probabilidade de $Y \geq 200$ e qual é a vida útil esperada do segundo componente "condicional em relação" a esse valor de X?

Questões desse tipo são respondidas pelo estudo das distribuições de probabilidades condicionais.

DEFINIÇÃO

Sejam X e Y duas vas contínuas com fdp conjunta $f(x, y)$ e fdp marginal de X, $f_X(x)$. Então, para qualquer valor x de X para o qual $f_X(x) > 0$, a **função de densidade de probabilidade condicional de Y dado que $X = x$** é

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} \qquad -\infty < y < \infty$$

Se X e Y forem discretas, substituir as fdps por fmps nesta definição fornecerá a **função de massa de probabilidade condicional de Y quando $X = x$.**

Observe que a definição de $f_{Y|X}(y \mid x)$ corresponde à de $P(B \mid A)$, a probabilidade condicional da ocorrência de B, dado que A tenha ocorrido. Uma vez que a fdp ou fmp condicional tiver sido determinada, questões do tipo propostas no início desta subseção podem ser respondidas pela integração ou pela soma de um conjunto apropriado de valores Y.

Exemplo 5.12

Reconsidere a situação dos Exemplos 5.3 e 5.4 que envolvem X = proporção de tempo em que o guichê de automóveis de um banco está ocupado e Y = proporção análoga para o guichê de atendimento a pessoas. A fdp condicional de Y dado que $X = 0{,}8$ é

$$f_{Y|X}(y \mid 0{,}8) = \frac{f(0{,}8, y)}{f_X(0{,}8)} = \frac{1{,}2(0{,}8 + y^2)}{1{,}2(0{,}8) + 0{,}4} = \frac{1}{34}(24 + 30y^2) \qquad 0 < y < 1$$

A probabilidade de o guichê de pessoas estar ocupado no máximo metade do tempo, dado que $X = 0{,}8$, é

$$P(Y \leq 0{,}5 \mid X = 0{,}8) = \int_{-\infty}^{0{,}5} f_{Y|X}(y \mid 0{,}8)\, dy = \int_{0}^{0{,}5} \frac{1}{34}(24 + 30y^2)\, dy = 0{,}390$$

Usar a fdp marginal de Y fornece $P(Y \leq 0{,}5) = 0{,}350$. Além disso, $E(Y) = 0{,}6$, enquanto a proporção esperada de tempo em que o guichê de pessoas está ocupado, dado que $X = 0{,}8$ (uma expectativa *condicional*) é

$$E(Y \mid X = 0{,}8) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid 0{,}8)\, dy = \frac{1}{34}\int_{0}^{1} y(24 + 30y^2)\, dy = 0{,}574 \quad ■$$

Exercícios | Seção 5.1 (1–21)

1. Um posto de gasolina tem ilhas de auto-serviço e de serviço completo. Em cada ilha, há uma única bomba de auto-serviço de gasolina comum com duas mangueiras. Sejam X = número de mangueiras em uso na ilha de auto-serviço em um momento específico e Y = número de mangueiras na ilha de serviço completo em uso naquele mesmo momento. A fmp de X e Y é mostrada na tabela a seguir:

$p(x, y)$		y = 0	1	2
	0	0,10	0,04	0,02
x	1	0,08	0,20	0,06
	2	0,06	0,14	0,30

a. Qual é $P(X = 1 \text{ e } Y = 1)$?

b. Calcule $P(X \leq 1 \text{ e } Y \leq 1)$.

c. Descreva o evento $\{X \neq 0 \text{ e } Y \neq 0\}$ e calcule sua probabilidade.

d. Calcule a fmp marginal de X e de Y. Usando $p_X(x)$, qual é $P(X \leq 1)$?

e. X e Y são vas independentes? Explique.

2. Quando um automóvel é parado por uma patrulha de segurança, os pneus são verificados quanto ao desgaste, assim como o alinhamento de cada farol. Seja X o número de faróis que precisam de ajuste e Y o número de pneus com defeito.

a. Se X e Y são independentes com $p_X(0) = 0{,}5$, $p_X(1) = 0{,}3$, $p_X(2) = 0{,}2$, e $p_Y(0) = 0{,}6$, $p_Y(1) = 0{,}1$, $p_Y(2) =$

$p_Y(3) = 0{,}05$, $p_Y(4) = 0{,}2$, mostre a fmp conjunta de (X, Y) em uma tabela de probabilidade conjunta.

b. Calcule $P(X \leq 1 \text{ e } Y \leq 1)$ pela tabela de probabilidade conjunta e verifique se ela é igual ao produto $P(X \leq 1) \cdot P(Y \leq 1)$.

c. Qual é $P(X + Y = 0)$ (a probabilidade de nenhuma violação)?

d. Calcule $P(X + Y \leq 1)$.

3. Um determinado mercado tem uma fila de caixa expressa e uma fila de caixa superexpressa. Represente por X_1 o número de clientes na fila da caixa expressa em um determinado horário do dia e por X_2 o número de clientes na fila da caixa superexpressa no mesmo horário. Suponha que a fmp conjunta de X_1 e X_2 seja dada na tabela a seguir.

		x_2: 0	1	2	3
x_1	0	0,08	0,07	0,04	0,00
	1	0,06	0,15	0,05	0,04
	2	0,05	0,04	0,10	0,06
	3	0,00	0,03	0,04	0,07
	4	0,00	0,01	0,05	0,06

a. Qual é $P(X_1 = 1, X_2 = 1)$, isto é, a probabilidade de haver exatamente um cliente em cada fila?

b. Qual é $P(X_1 = X_2)$, isto é, a probabilidade de os números de clientes nas duas filas serem iguais?

c. Represente por A o evento de haver no mínimo dois clientes a mais em uma fila do que na outra. Expresse A em termos de X_1 e X_2, e calcule a probabilidade desse evento.

d. Qual é a probabilidade de o número total de clientes nas duas filas ser exatamente quatro? E no mínimo quatro?

4. Retorne à situação descrita no Exercício 3.

a. Determine a fmp marginal de X_1 e, depois, calcule o número esperado de clientes na fila da caixa expressa.

b. Determine a fmp marginal de X_2.

c. Pela inspeção das probabilidades $P(X_1 = 4)$, $P(X_2 = 0)$, e $P(X_1 = 4, X_2 = 0)$, X_1 e X_2 são variáveis aleatórias independentes? Explique.

5. O número de clientes esperando no serviço de embalagens para presente em uma loja é uma va X com valores possíveis 0, 1, 2, 3, 4 e probabilidades correspondentes 0,1, 0,2, 0,3, 0,25, 0,15. Um cliente selecionado aleatoriamente terá 1, 2 ou 3 pacotes para embrulhar com probabilidades 0,6, 0,3 e 0,1, respectivamente. Seja Y = o número total de pacotes a serem embrulhados para os clientes que estão na fila (assuma que o número de pacotes de um cliente seja independente do número de pacotes de qualquer outro cliente).

a. Determine $P(X = 3, Y = 3)$, isto é, $p(3, 3)$.

b. Determine $p(4, 11)$.

6. Seja X número de câmeras digitais Canon vendidas durante uma determinada semana por uma loja. A fmp de X é

x	0	1	2	3	4
$p_X(x)$	0,1	0,2	0,3	0,25	0,15

Sessenta por cento de todos os clientes que compram essas câmeras também compram uma garantia estendida. Seja Y o número de compradores que durante essa semana adquiriram garantia estendida.

a. Qual é $P(X = 4, Y = 2)$? [*Sugestão*: essa probabilidade é igual a $P(Y = 2 \mid X = 4) \cdot P(X = 4)$; imagine agora que as quatro compras sejam quatro tentativas de um experimento binomial, com sucesso em uma tentativa correspondente à aquisição de uma garantia estendida].

b. Calcule $P(X = Y)$.

c. Determine a fmp conjunta de X e Y e, depois, a fmp marginal de Y.

7. A distribuição de probabilidades conjuntas do número X de carros e do número Y de ônibus por ciclo de semáforo em uma conversão à esquerda proposta é exibida na tabela de probabilidade conjunta a seguir.

$p(x, y)$		y: 0	1	2
x	0	0,025	0,015	0,010
	1	0,050	0,030	0,020
	2	0,125	0,075	0,050
	3	0,150	0,090	0,060
	4	0,100	0,060	0,040
	5	0,050	0,030	0,020

a. Qual é a probabilidade de haver exatamente um carro e exatamente um ônibus em um ciclo?

b. Qual é a probabilidade de haver no máximo um carro e no máximo um ônibus em um ciclo?

c. Qual é a probabilidade de haver exatamente um carro em um ciclo? E exatamente um ônibus?

d. Suponha que uma pista de conversão à esquerda tenha capacidade para cinco carros e um ônibus, equivalente a três carros. Qual é a probabilidade de excesso de fluxo em um ciclo?

e. X e Y são vas independentes? Explique.

8. Um armazém tem atualmente 30 componentes de certo tipo, dos quais 8 são fornecidos pelo fornecedor 1, 10 pelo fornecedor 2 e 12 pelo fornecedor 3. Seis deles serão aleatoriamente selecionados para uma montagem. Seja X = número de componentes selecionados do fornecedor 1, Y = número de componentes selecionados do fornecedor 2 e $p(x, y)$ a fmp conjunta de X e Y.

a. Qual é $p(3, 2)$? [*Sugestão*: cada amostra de tamanho 6 tem a mesma probabilidade de ser selecionada. Portanto, $p(3, 2)$ = (número de resultados com $X = 3$ e $Y = 2$)/(número total de resultados). Use a regra do produto para contagem a fim de obter o numerador e o denominador].

b. Usando a lógica do item (a), obtenha $p(x, y)$. (Isso pode ser visto como uma distribuição hipergeométrica multivariada – amostragem sem reposição de

uma população finita consistindo de mais de duas categorias).

9. Admite-se que cada pneu dianteiro de um determinado tipo de veículo deve ter a pressão de 26 psi. Suponha que a pressão real em cada pneu seja uma variável aleatória X para o pneu direito e Y para o esquerdo, com fdp conjunta

$$f(x, y) = \begin{cases} K(x^2 + y^2) & 20 \le x \le 30, 20 \le y \le 30 \\ 0 & \text{caso contrário} \end{cases}$$

a. Qual é o valor de K?
b. Qual é a probabilidade de os dois pneus estarem com pressão inferior à ideal?
c. Qual é a probabilidade de a diferença de pressão de ar entre os dois pneus ser no máximo 2 psi?
d. Determine a distribuição (marginal) da pressão de ar só do pneu direito.
e. X e Y são vas independentes?

10. Annie e Alvie combinaram de se encontrar entre 17 e 18 horas para jantar em um restaurante local de comida saudável. Seja X = a hora de chegada de Annie e Y = a de Alvie. Suponha que X e Y sejam independentes com distribuição uniforme no intervalo [5, 6].
a. Qual é a fdp conjunta de X e Y?
b. Qual é a probabilidade de ambos chegarem entre 17h15 e 17h45?
c. Se o primeiro a chegar esperar apenas 10 minutos antes de ir comer em outro lugar, qual é a probabilidade de eles comerem no restaurante de comida saudável? [*Sugestão*: o evento de interesse é $A = \{(x, y): |x - y| \le \frac{1}{6}\}$.].

11. Dois professores diferentes enviaram suas provas finais para tiragem de cópias. Seja X o número de erros tipográficos na prova do primeiro professor e Y o número de tais erros para o segundo professor. Suponha que X tenha distribuição de Poisson com parâmetro λ, Y tenha distribuição de Poisson com parâmetro θ e que X e Y sejam independentes.
a. Qual é a fmp conjunta de X e Y?
b. Qual é a probabilidade de no máximo um erro ser cometido em ambos os exames combinados?
c. Obtenha a expressão geral para a probabilidade de o número total de erros nos dois exames ser m (onde m é um inteiro não-negativo).[*Sugestão*: $A = \{(x, y): x + y = m\} = \{(m, 0), (m - 1, 1),..., (1, m - 1), (0, m)\}$. Agora some a fmp conjunta em $(x, y) \in A$ e use o teorema binomial, que diz

$$\sum_{k=0}^{m} \binom{m}{k} a^k b^{m-k} = (a + b)^m$$

para qualquer a, b.]

12. Dois componentes de um minicomputador possuem a fdp conjunta a seguir para seus tempos de vida útil X e Y.

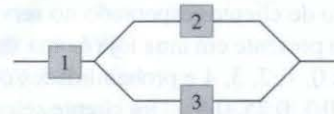
$$f(x, y) = \begin{cases} xe^{-x(1+y)} & x \ge 0 \text{ e } y \ge 0 \\ 0 & \text{caso contrário} \end{cases}$$

a. Qual é a probabilidade da vida útil X do primeiro componente exceder 3?
b. Quais são as fdps marginais de X e Y? Os dois períodos de vida útil são independentes? Explique.
c. Qual é a probabilidade de o tempo de vida de pelo menos um componente exceder 3?

13. Você tem duas lâmpadas para uma determinada luminária. Seja X = vida útil da primeira lâmpada e Y = vida útil da segunda lâmpada (ambas de mil horas). Suponha que X e Y sejam independentes e que cada uma tenha distribuição exponencial com parâmetro $\lambda = 1$.
a. Qual é a fdp conjunta de X e Y?
b. Qual é a probabilidade de cada lâmpada durar no máximo 1000 horas (isso é, $X \le 1$ e $Y \le 1$)?
c. Qual é a probabilidade de a vida útil total das duas lâmpadas ser no máximo 2? [*Sugestão*: desenhe a região $A = \{(x, y): x \ge 0, y \ge 0, x + y \le 2\}$ antes da integração.]
d. Qual é a probabilidade de a vida útil total estar entre 1 e 2?

14. Suponha que você tenha 10 lâmpadas, que a vida útil de cada uma seja independente das demais e que cada vida útil tenha distribuição exponencial com parâmetro λ.
a. Qual é a probabilidade de as 10 lâmpadas falharem antes do tempo t?
b. Qual é a probabilidade de exatamente k das 10 lâmpadas falharem antes do tempo t?
c. Suponha que nove lâmpadas tenham vida útil exponencialmente distribuída com parâmetro λ e que a lâmpada que sobrou tenha vida útil exponencialmente distribuída com parâmetro θ (é feita por outro fabricante). Qual é a probabilidade de exatamente cinco das 10 lâmpadas falharem antes do tempo t?

15. Considere um sistema formado por três componentes, conforme ilustração. O sistema continuará funcionando desde que o primeiro componente funcione e que o componente 2 ou o 3 funcionem também. Represente por X_1, X_2 e X_3 os tempos de vida dos componentes 1, 2 e 3, respectivamente. Suponha que os X_is sejam independentes entre si e que cada X_i possua uma distribuição exponencial com parâmetro λ.



a. Seja Y = tempo de vida do sistema. Obtenha a função de distribuição acumulada de Y e calcule a diferença para obter a fdp. [*Sugestão*: $F(y) = P(Y \le y)$; expresse o evento $\{Y \le y\}$ em termos de uniões e/ou interseções dos três eventos $\{X_1 \le y\}$, $\{X_2 \le y\}$, e $\{X_3 \le y\}$.]
b. Calcule a vida útil esperada do sistema.

16. a. Para $f(x_1, x_2, x_3)$, como dado no Exemplo 5.10, calcule a **função de densidade marginal conjunta** de X_1 e somente de X_3 (integrando em x_2).

b. Qual é a probabilidade de que as rochas dos tipos 1 e 3 juntas formem no máximo 50% da amostra? [*Sugestão*: use o resultado do item (a).]

c. Calcule a fdp marginal apenas de X_1. [*Sugestão*: use o resultado do item (a).]

17. Um ecologista deseja selecionar um ponto dentro de uma região circular de amostragem, de acordo com uma distribuição uniforme (na prática, isso poderia ser feito selecionando-se primeiro uma direção e, depois, uma distância do centro naquela direção). Seja X = coordenada x do ponto selecionado e Y = coordenada y desse ponto. Se o círculo tiver centro em (0, 0) e raio R, então a fdp conjunta de X e Y é

$$f(x, y) = \begin{cases} \dfrac{1}{\pi R^2} & x^2 + y^2 \leq R^2 \\ 0 & \text{caso contrário} \end{cases}$$

a. Qual é a probabilidade de o ponto selecionado estar dentro de $R/2$ do centro da região circular? [*Sugestão*: desenhe a região de densidade positiva D. Já que $f(x, y)$ é constante em D, calcular a probabilidade resume-se em calcular uma área.]

b. Qual é a probabilidade de X e Y serem diferentes de 0 no máximo $R/2$?

c. Responda ao item (b) com $R/\sqrt{2}$ substituindo $R/2$.

d. Qual é a fdp marginal de X? De Y? X e Y são independentes?

18. Refira-se ao Exercício 1 e responda as seguintes perguntas:

a. Dado que $X = 1$, determine a fmp condicional de Y, isto é, $p_{Y|X}(0 \mid 1)$, $p_{Y|X}(1 \mid 1)$, e $p_{Y|X}(2 \mid 1)$.

b. Dado que duas mangueiras estão sendo utilizadas na ilha de auto-serviço, qual é a fmp condicional do número de mangueiras em uso na ilha de serviço completo?

c. Use o resultado do item (b) para calcular a probabilidade condicional $P(Y \leq 1 \mid X = 2)$.

d. Dado que duas mangueiras estão sendo utilizadas na ilha de serviço completo, qual é a fmp condicional do número em uso na ilha de auto-serviço?

19. A fdp conjunta de pressões dos pneus dianteiros é dada no Exercício 9.

a. Determine a fdp condicional de Y, dado que $X = x$, e a fdp condicional de X, dado que $Y = y$.

b. Se a pressão no pneu direito for 22 psi, qual é a probabilidade de o pneu esquerdo ter uma pressão de no mínimo 25 psi? Compare com $P(Y \geq 25)$.

c. Se a pressão no pneu direito for 22 psi, qual é a pressão esperada no pneu esquerdo e qual é o desvio padrão da pressão desse pneu?

20. Sejam X_1, X_2, e X_3 as vidas úteis dos componentes 1, 2 e 3 em um sistema de três componentes.

a. Como você definiria a fdp condicional de X_3, dado que $X_1 = x_1$ e $X_2 = x_2$?

b. Como você definiria a fdp condicional conjunta de X_2 e X_3 dado que $X_1 = x_1$?

21. Que condição de $f_{Y|X}(y \mid x)$ equivale à independência de X e Y?

5.2 Valores Esperados, Covariância e Correlação

Vimos anteriormente que qualquer função $h(X)$ de uma única va X é por si mesma uma variável aleatória. Entretanto, para calcular $E[h(X)]$, não foi necessário obter a distribuição de probabilidades de $h(X)$; ao contrário, $E[h(X)]$ foi calculado como uma média ponderada dos valores de $h(x)$, em que a função do peso foi a fmp $p(x)$ ou a fdp $f(x)$ de X. Um resultado semelhante é obtido para a função $h(X, Y)$ de duas variáveis aleatórias de distribuição conjunta.

PROPOSIÇÃO Sejam X e Y as vas de distribuição conjunta com fmp $p(x, y)$ ou fdp $f(x, y)$, conforme as variáveis sejam discretas ou contínuas. Então, o valor esperado de uma função $h(X, Y)$, representada por $E[h(X, Y)]$ ou $\mu_{h(X,Y)}$, é dado por

$$E[h(X, Y)] = \begin{cases} \displaystyle\sum_x \sum_y h(x, y) \cdot p(x, y) & \text{se } X \text{ e } Y \text{ forem discretos} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y) \cdot f(x, y)\, dx\, dy & \text{se } X \text{ e } Y \text{ forem contínuos} \end{cases}$$

Exemplo 5.13

Cinco amigos compraram ingressos para um determinado show. Se os ingressos forem dos lugares de 1-5 em certa fileira, e os ingressos forem distribuídos aleatoriamente entre os cinco, qual é o número esperado de assentos que separam quaisquer dois dos cinco? Sejam X e Y os números dos lugares da primeira e da segunda pessoas, respectivamente. Possíveis pares de (X, Y) são $\{(1, 2), (1, 3),..., (5, 4)\}$, e a fmp conjunta de (X, Y) é

$$p(x, y) = \begin{cases} \frac{1}{20} & x = 1, \ldots, 5; y = 1, \ldots, 5; x \neq y \\ 0 & \text{caso contrário} \end{cases}$$

O número de lugares que separam as duas pessoas é $h(X, Y) = |X - Y| - 1$. A tabela a seguir fornece $h(x, y)$ para cada par (x, y) possível.

$h(x, y)$		x = 1	2	3	4	5
	1	—	0	1	2	3
	2	0	—	0	1	2
y	3	1	0	—	0	1
	4	2	1	0	—	0
	5	3	2	1	0	—

Dessa forma,

$$E[h(X, Y)] = \sum_{(x, y)}\sum h(x, y) \cdot p(x, y) = \sum_{x=1}^{5} \sum_{\substack{y=1 \\ x \neq y}}^{5} (|x - y| - 1) \cdot \frac{1}{20} = 1$$

Exemplo 5.14

No Exemplo 5.5, a fdp conjunta da quantidade X de amêndoas e da quantidade Y de castanhas de caju em uma lata de castanhas de 1 libra foi

$$f(x, y) = \begin{cases} 24xy & 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Se 1 libra de amêndoas custar à empresa US\$ 1,00, 1 libra de castanhas de caju custar US\$ 1,50 e 1 libra de amendoins custar US\$ 0,50, então o custo total do conteúdo de uma lata é

$$h(X, Y) = (1)X + (1,5)Y + (0,5)(1 - X - Y) = 0,5 + 0,5X + Y$$

(uma vez que $1 - X - Y$ do peso consiste em amendoins). O custo total esperado é

$$E[h(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y) \cdot f(x, y)\, dx\, dy$$

$$= \int_0^1\int_0^{1-x} (0,5 + 0,5x + y) \cdot 24xy\, dy\, dx = \text{US\$ } 1,10$$ ■

O método para calcular o valor esperado de uma função $h(X_1,..., X_n)$ *de n* variáveis aleatórias é semelhante ao método usado para duas variáveis aleatórias. Se os X_is forem discretos, $E[h(X_1,..., X_n)]$ é uma soma n-dimensional; se os X_is forem contínuos, é uma integral n-dimensional.

Covariância

Quando duas variáveis aleatórias X e Y não são independentes, geralmente é de interesse avaliar quão fortemente estão relacionadas uma com a outra.

DEFINIÇÃO A **covariância** entre duas vas X e Y é

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

$$= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y) & X, Y \text{ discretas} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)\, dx\, dy & X, Y \text{ contínuas} \end{cases}$$

O fundamento lógico da definição é o seguinte. Suponha que X e Y tenham uma relação positiva forte entre si, pela qual queremos dizer que valores grandes de X tendem a ocorrer com valores grandes de Y, e valores pequenos de X, com valores pequenos de Y. Então, a maior parte da massa ou densidade de probabilidade estará associada a $(x - \mu_X)$ e $(y - \mu_Y)$, ambos positivos (X e Y acima de suas respectivas médias) ou ambos negativos, de modo que o produto $(x - \mu_X)(y - \mu_Y)$ tende a ser positivo. Dessa forma, para uma relação positiva forte, Cov(X, Y) deve ser positiva. Para uma relação negativa forte, os sinais de $(x - \mu_X)$ e $(y - \mu_Y)$ tenderão a ser opostos, levando a um produto negativo. Assim, para uma relação negativa forte, Cov(X, Y) deve ser negativa. Se X e Y não estão fortemente relacionadas, os produtos positivo e negativo tenderão a cancelar um ao outro, produzindo uma covariância próxima de 0. A Figura 5.4 ilustra as diferentes possibilidades. A covariância depende tanto do conjunto de pares possíveis como das probabilidades. Na Figura 5.4, as probabilidades poderiam ser trocadas sem alterar o conjunto de pares possíveis, o que pode mudar drasticamente o valor da Cov(X, Y).

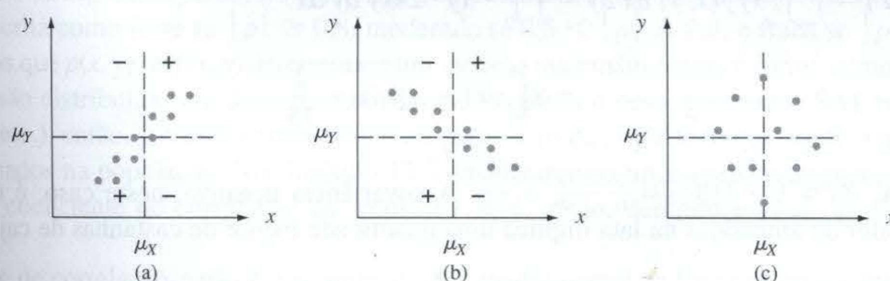


Figura 5.4 $p(x, y) = 1/10$ para cada um dos 10 pares correspondentes aos pontos indicados, (a) covariância positiva; (b) covariância negativa; (c) covariância próxima de zero

Exemplo 5.15

As fmps marginal e conjunta de X = quantidade dedutível das apólices de seguro de automóvel e Y = quantidade dedutível das apólices de seguro residencial no Exemplo 5.1 eram

$p(x, y)$		y = 0	100	200
x	100	0,20	0,10	0,20
	250	0,05	0,15	0,30

x	100	250
$p_X(x)$	0,5	0,5

y	0	100	200
$p_Y(y)$	0,25	0,25	0,5

onde $\mu_X = \sum x p_X(x) = 175$ e $\mu_Y = 125$. Portanto,

$$\begin{aligned}\text{Cov}(X, Y) &= \sum_{(x,y)}\sum (x - 175)(y - 125)p(x, y) \\ &= (100 - 175)(0 - 125)(0{,}20) + \cdots \\ &\quad + (250 - 175)(200 - 125)(0{,}30) \\ &= 1875\end{aligned}$$

■

A fórmula simplificada a seguir para Cov(*X*, *Y*) facilita os cálculos.

PROPOSIÇÃO

$$\text{Cov}(X, Y) = E(XY) - \mu_X \cdot \mu_Y$$

De acordo com essa fórmula, não é necessária nenhuma subtração intermediária; somente no final do cálculo $\mu_X \cdot \mu_Y$ é subtraído de $E(XY)$. A demonstração envolve expandir $(X - \mu_X)(Y - \mu_Y)$ e depois calcular o valor esperado de cada termo separadamente. Observe que $\text{Cov}(X, X) = E(X^2) - \mu_X^2 = V(X)$.

Exemplo 5.16 (continuação do Exemplo 5.5)

As fdps marginal e conjunta de X = quantidade de amêndoas e de Y = quantidade de castanhas de caju eram

$$f(x, y) = \begin{cases} 24xy & 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

$$f_X(x) = \begin{cases} 12x(1 - x)^2 & 0 \leq x \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

$f_Y(y)$ é obtido, substituindo x por y em $f_X(x)$. É fácil demonstrar que $\mu_X = \mu_Y = \frac{2}{5}$, e

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x, y)\, dx\, dy = \int_0^1\int_0^{1-x} xy \cdot 24xy\, dy\, dx$$

$$= 8\int_0^1 x^2(1 - x)^3\, dx = \frac{2}{15}$$

Assim, $\text{Cov}(X, Y) = \frac{2}{15} - (\frac{2}{5})(\frac{2}{5}) = \frac{2}{15} - \frac{4}{25} = -\frac{2}{75}$. A covariância negativa, neste caso, é razoável, porque uma quantidade maior de amêndoas na lata implica uma quantidade menor de castanhas de caju. ■

Pode parecer que a relação no exemplo do seguro seja bastante forte, uma vez que $\text{Cov}(X, Y) = 1875$, enquanto $\text{Cov}(X, Y) = -\frac{2}{75}$, no exemplo das castanhas, parece uma relação fraca. Infelizmente, a covariância possui uma deficiência grave que impossibilita a interpretação de um valor calculado. No exemplo do seguro, suponha que tivéssemos expressado a quantia dedutível em centavos e não em dólares. Então, $100X$ substituiria X, $100Y$ substituiria Y, e a covariância resultante seria $\text{Cov}(100X, 100Y) = (100)(100)\text{Cov}(X, Y) = 18.750.000$. Se, por outro lado, a quantia dedutível tivesse sido expressa em centenas de dólares, a covariância calculada teria sido $(0{,}01)(0{,}01)(1875) = 0{,}1875$. *A deficiência da covariância é que seu valor calculado depende diretamente das unidades de medida.* Teoricamente, a escolha das unidades não deve afetar a medida da intensidade da relação. O que é obtido definindo-se a escala da covariância.

Correlação

DEFINIÇÃO O **coeficiente de correlação** de X e Y, representado por $\text{Corr}(X, Y)$, $\rho_{X,Y}$, ou simplesmente ρ, é definido por

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Exemplo 5.17

É fácil verificar que no problema do seguro, do Exemplo 5.15, $E(X^2) = 36.250$, $\sigma_X^2 = 36.250 - (175)^2 = 5625$, $\sigma_X = 75$, $E(Y^2) = 22.500$, $\sigma_Y^2 = 6875$, e $\sigma_Y = 82{,}92$. O que dá

$$\rho = \frac{1875}{(75)(82{,}92)} = 0{,}301$$

■

A proposição a seguir mostra que ρ corrige a deficiência da $\text{Cov}(X, Y)$, além de sugerir como reconhecer a existência de uma relação (linear) forte.

PROPOSIÇÃO

1. Se a e c são ambos positivos ou negativos,

$$\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$$

2. Para quaisquer duas vas X e Y, $-1 \le \text{Corr}(X, Y) \le 1$.

A proposição 1 diz precisamente que o coeficiente de correlação não é afetado por mudança linear das unidades de medida (se, por exemplo, X = temperatura em °C, então, $9X/5 + 32$ = temperatura em °F). De acordo com a proposição 2, a relação positiva mais forte possível é evidenciada por $\rho = +1$, enquanto a relação negativa mais forte possível corresponde a $\rho = -1$. A demonstração da primeira proposição está esquematizada no Exercício 35, e a da segunda aparece no Exercício Suplementar 87, no final do capítulo. Para fins descritivos, a relação será descrita como forte se $|\rho| \ge 0{,}8$, moderada se $0{,}5 < |\rho| < 0{,}8$, e fraca se $|\rho| \le 0{,}5$.

Se pensarmos que $p(x, y)$ ou $f(x, y)$ determinam um modelo matemático para a forma como as duas variáveis numéricas X e Y são distribuídas em determinada população (altura e peso, pontuação SAT verbal e pontuação SAT quantitativa etc.), então, ρ é uma característica ou parâmetro da população que mede o quão fortemente X e Y estão relacionados na população. No Capítulo 12, consideraremos uma amostra de pares $(x_1, y_1),..., (x_n, y_n)$ da população. O coeficiente de correlação da amostra r será, então, definido e usado para fazer inferências sobre ρ.

O coeficiente de correlação ρ não é, na verdade, uma medida geral da força de uma relação.

PROPOSIÇÃO

1. Se X e Y são independentes, então $\rho = 0$, porém $\rho = 0$ não implica independência.
2. $\rho = 1$ ou -1 se e somente se $Y = aX + b$ para quaisquer números a e b com $a \neq 0$.

Essa proposição diz que ρ é uma medida do grau da relação **linear** entre X e Y, e somente quando as duas variáveis estiverem perfeitamente relacionadas de forma linear é que ρ assumirá os valores extremos positivo ou negativo. Um ρ menor que 1 em valor absoluto indica somente que a relação não é completamente linear, mas que ainda pode haver uma relação não-linear bastante forte. Além disso, $\rho = 0$ não implica que X e Y sejam independentes, mas apenas que há ausência completa de relação linear. Quando $\rho = 0$, X e Y são ditos **não-correlacionados**. Duas variáveis podem ser não-correlacionadas, porém altamente dependentes, pois pode existir uma relação não-linear forte; assim, cuidado para não tirar muitas conclusões a partir de $\rho = 0$.

Exemplo 5.18

Sejam X e Y vas discretas com fmp conjunta

$$p(x, y) = \begin{cases} \dfrac{1}{4} & (x, y) = (-4, 1), (4, -1), (2, 2), (-2, -2) \\ 0 & \text{caso contrário} \end{cases}$$

Os pontos que recebem massa de probabilidade positiva estão identificados no sistema de coordenadas (*x*, *y*) na Figura 5.5. É evidente, pela figura, que o valor de *X* é determinado completamente pelo valor de *Y* e vice-versa, de modo que as duas variáveis são totalmente dependentes. Entretanto, pela simetria $\mu_X = \mu_Y = 0$ e $E(XY) = (-4)\frac{1}{4} + (-4)\frac{1}{4} + (4)\frac{1}{4} + (4)\frac{1}{4} = 0$, assim, $\text{Cov}(X, Y) = E(XY) - \mu_X \cdot \mu_Y = 0$ e dessa forma $\rho_{X,Y} = 0$. Embora haja perfeita dependência, também há ausência completa de qualquer relação linear!

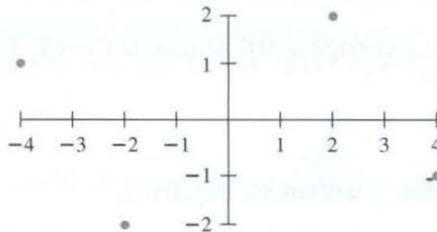


Figura 5.5 A população de pares do Exemplo 5.18 ■

Um valor de ρ próximo de 1 não implica necessariamente que aumentar o valor de *X cause um aumento em Y*. Implica somente que valores grandes de *X* estão *associados* a valores grandes de *Y*. Por exemplo: na população de crianças, o tamanho do vocabulário e o número de cáries são correlacionados de forma bastante positiva, mas certamente não é verdade que as cáries façam o vocabulário aumentar. Ao contrário, os valores de ambas as variáveis tendem a aumentar com a idade das crianças, uma terceira variável. Para as crianças de certa idade, provavelmente haja correlação muito baixa entre o número de cáries e o tamanho do vocabulário. Resumindo, associação (uma alta correlação) não é a mesma coisa que causa.

Exercícios | Seção 5.2 (22–36)

22. Um instrutor fez perguntas curtas que consistem de duas partes. Para um estudante selecionado aleatoriamente, seja *X* = o número de pontos ganhos na primeira parte e *Y* = o número de pontos ganhos na segunda. Suponha que a fmp conjunta de *X* e *Y* seja dada na tabela a seguir.

	(x, y)	y = 0	5	10	15
x	0	0,02	0,06	0,02	0,10
	5	0,04	0,15	0,20	0,10
	10	0,01	0,15	0,14	0,01

a. Se a pontuação constante do boletim for o número total de pontos ganhos nas duas partes, qual é a pontuação registrada esperada $E(X + Y)$?

b. Se o máximo das duas pontuações for registrado, qual é a pontuação registrada esperada?

23. A diferença entre o número de clientes na fila da caixa expressa e o número de clientes na caixa superexpressa no Exercício 3 é $X_1 - X_2$. Calcule a diferença esperada.

24. Seis indivíduos, incluindo A e B, sentam-se ao redor de uma mesa redonda de forma totalmente aleatória. Suponha que os lugares sejam numerados de 1, ..., 6. Seja *X* = o número do assento de A e *Y* = o número do assento de B. Se A mandar um bilhete, ao redor da mesa, para B, na direção em que estão mais próximos, quantos indivíduos (incluindo A e B) você espera que peguem o bilhete?

25. Um agrimensor deseja traçar uma região quadrada cujos lados têm comprimento *L*. Entretanto, devido a erro de medição, ele esboça um retângulo, em que os lados norte e sul têm comprimento *X*, e os lados leste e oeste, comprimento *Y*. Suponha que *X* e *Y* sejam independentes e que cada um seja distribuído uniformemente no intervalo $[L - A, L + A]$ (onde $0 < A < L$). Qual é a área esperada do retângulo resultante?

26. Considere uma pequena balsa que possa acomodar carros e ônibus. O pedágio para carros são US$ 3, e para ônibus, US$ 10. Sejam *X* e *Y* o número de carros e ônibus, respectivamente, levado em uma única viagem. Suponha que a distribuição conjunta de *X* e *Y* seja aquela dada na tabela do Exercício 7. Calcule a receita esperada de uma única viagem.

27. Annie e Alvie combinaram encontrar-se para almoçar juntos entre meio-dia e 13h. Represente por *X* a hora de chegada de Annie, por *Y* a de Alvie, e suponha que *X* e *Y* sejam independentes com fdps

$$f_X(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & \text{caso contrário} \end{cases}$$

$$f_Y(y) = \begin{cases} 2y & 0 \leq y \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Qual é o tempo esperado que a pessoa que chega primeiro deve aguardar pela outra? [*Sugestão*: $h(X, Y) = |X - Y|$.]

28. Mostre que, se X e Y são vas independentes, então $E(XY) = E(X) \cdot E(Y)$. Aplique essa expressão no Exercício 25. [*Sugestão*: considere o caso contínuo com $f(x, y) = f_X(x) \cdot f_Y(y)$.]

29. Calcule o coeficiente de correlação ρ de X e Y do Exemplo 5.16 (a covariância já foi calculada).

30. a. Calcule a covariância de X e Y no Exercício 22.
b. Calcule ρ de X e Y no mesmo exercício.

31. a. Calcule a covariância entre X e Y no Exercício 9.
b. Calcule o coeficiente de correlação ρ para X e Y.

32. Reconsidere as vidas úteis X e Y do componente do minicomputador, como descrito no Exercício 12. Determine $E(XY)$. O que se pode dizer sobre $\text{Cov}(X, Y)$ e ρ?

33. Use o resultado do Exercício 28 para mostrar que, quando X e Y são independentes, $\text{Cov}(X, Y) = \text{Corr}(X, Y) = 0$.

34. a. Lembrando-se da definição de σ^2 para uma única va X, escreva uma fórmula apropriada para calcular a variância da função $h(X, Y)$ de duas variáveis aleatórias. [*Sugestão*: lembre-se de que a variância é apenas um valor esperado especial.]
b. Use essa fórmula para calcular a variância da pontuação registrada $h(X, Y)$ [$= \max(X, Y)$] na parte (b) do Exercício 22.

35. a. Use as regras de valor esperado para demonstrar que $\text{Cov}(aX + b, cY + d) = ac\,\text{Cov}(X, Y)$.
b. Use a parte (a) junto com as regras de variância e desvio padrão para demonstrar que $\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$ quando a e c têm o mesmo sinal.
c. O que acontece se a e c tiverem sinais contrários?

36. Mostre que, se $Y = aX + b$ ($a \neq 0$), então $\text{Corr}(X, Y) = +1$ ou -1. Sob quais condições $\rho = +1$?

5.3 Estatísticas e Suas Distribuições

As observações de uma única amostra foram denotadas no Capítulo 1 por $x_1, x_2, \ldots, x_n$. Considere a seleção de duas amostras diferentes de tamanho n oriundas da mesma distribuição da população. Os x_is da segunda amostra serão sempre virtualmente diferentes, pelo menos um pouco, dos x_is da primeira amostra. Por exemplo: uma primeira amostra de $n = 3$ carros de um tipo específico pode resultar em consumos de combustível $x_1 = 30{,}7$, $x_2 = 29{,}4$, $x_3 = 31{,}1$, enquanto uma segunda amostra pode fornecer $x_1 = 28{,}8$, $x_2 = 30{,}0$ e $x_3 = 31{,}1$. Antes de obtermos os dados, há incerteza acerca do valor de cada x_i. Devido a essa incerteza, *antes* de os dados estarem disponíveis, consideramos cada observação como uma variável aleatória e denotamos a amostra por $X_1, X_2, \ldots, X_n$ (letras maiúsculas para variáveis aleatórias).

Essa variação nos valores observados implica, por sua vez, que o valor de qualquer função das observações da amostra – como a sua média, o seu desvio padrão ou a sua quarta dispersão – também varia de amostra para amostra. Isto é, antes de obter $x_1, \ldots, x_n$, há incerteza quanto ao valor de $\bar{x}$, o valor de s, e assim por diante.

Exemplo 5.19

Suponha que a resistência do material de um tipo específico de espécime selecionado aleatoriamente possui uma distribuição de Weibull com valores de parâmetro $\alpha = 2$ (formato) e $\beta = 5$ (escala). A curva de densidade correspondente é exibida na Figura 5.6. As fórmulas da Seção 4.5 fornecem

$$\mu = E(X) = 4{,}4311 \qquad \tilde{\mu} = 4{,}1628 \qquad \sigma^2 = V(X) = 5{,}365 \qquad \sigma = 2{,}316$$

A média excede a mediana devido à inclinação positiva da distribuição.

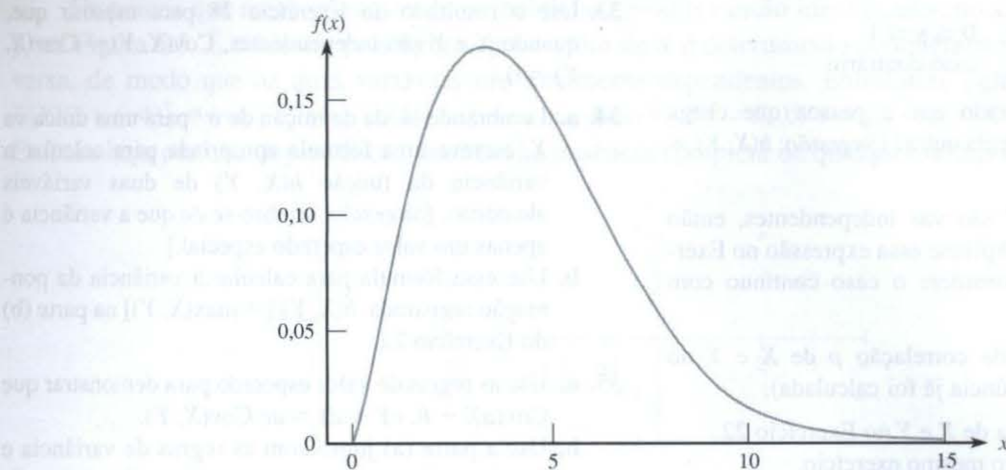


Figura 5.6 A curva de densidade de Weibull do Exemplo 5.19

Usamos MINITAB para gerar seis amostras diferentes, cada uma com $n = 10$ a partir dessa distribuição (resistências do material para seis grupos diferentes de 10 espécimes cada). Os resultados são exibidos na Tabela 5.1, seguidos pelos valores da média, mediana e desvio padrão de cada uma das amostras. Repare primeiro que as 10 observações em qualquer amostra específica são diferentes das observações em qualquer outra amostra. Em segundo lugar, os seis valores da média amostral são diferentes uns dos outros, como os seis valores da mediana e do desvio padrão da amostra. O mesmo acontece com as médias trincadas de 10% da amostra, com suas quartas dispersões da amostra e assim por diante.

Tabela 5.1 Amostras da distribuição de Weibull do Exemplo 5.19

Amostra	1	2	3	4	5	6
1	6,1171	5,07611	3,46710	1,55601	3,12372	8,93795
2	4,1600	6,79279	2,71938	4,56941	6,09685	3,92487
3	3,1950	4,43259	5,88129	4,79870	3,41181	8,76202
4	0,6694	8,55752	5,14915	2,49759	1,65409	7,05569
5	1,8552	6,82487	4,99635	2,33267	2,29512	2,30932
6	5,2316	7,39958	5,86887	4,01295	2,12583	5,94195
7	2,7609	2,14755	6,05918	9,08845	3,20938	6,74166
8	10,2185	8,50628	1,80119	3,25728	3,23209	1,75468
9	5,2438	5,49510	4,21994	3,70132	6,84426	4,91827
10	4,5590	4,04525	2,12934	5,50134	4,20694	7,26081
$\bar{x}$	4,401	5,928	4,229	4,132	3,620	5,761
$\tilde{x}$	4,360	6,144	4,608	3,857	3,221	6,342
s	2,642	2,062	1,611	2,124	1,678	2.496

Além disso, o valor da média amostral de qualquer amostra específica pode ser considerado uma *estimativa por pontos* ("ponto" porque é um único número, correspondente a um único ponto na linha dos números) da média da população μ, cujo valor conhecido é 4,4311. Nenhuma das estimativas dessas seis amostras é idêntica à que está sendo estimada. As estimativas da segunda e da sexta amostras são muito grandes, enquanto a quinta amostra fornece uma subestimativa substancial. De forma semelhante, o desvio padrão da amostra fornece uma estimativa por ponto do desvio padrão da população. As seis estimativas resultantes estão erradas em pelo menos uma pequena quantidade.

Resumindo, os valores das observações individuais variam de amostra para amostra. Dessa forma, geralmente, o valor de qualquer quantidade calculada a partir dos dados amostrais e o valor de uma amostra característica usada como estimativa da característica da população correspondente, virtualmente, jamais coincidirão com o que está sendo estimado. ■

DEFINIÇÃO **Estatística** é qualquer quantidade cujo valor possa ser calculado com base nos dados da amostra. Antes de obter os dados, existe a incerteza quanto ao valor que resultará de qualquer estatística específica. Portanto, estatística é uma variável aleatória e será representada por uma letra maiúscula; uma letra minúscula será usada para representar o valor calculado ou observado da estatística.

Assim, a média amostral, considerada como estatística (antes de a amostra ser selecionada ou de o experimento ter sido feito), é representada por $\overline{X}$; o valor calculado dessa estatística é $\overline{x}$. De forma semelhante, S representa o desvio padrão da amostra visto como estatística e seu valor calculado é s. Se amostras de dois tipos diferentes de tijolos forem selecionadas e as resistências à compressão individuais forem denotadas por $X_1,\ldots, X_m$ e $Y_1,\ldots, Y_n$, respectivamente, então a estatística $\overline{X} - \overline{Y}$, diferença entre as duas resistências à compressão média amostral, geralmente é de grande interesse.

Qualquer estatística, sendo uma variável aleatória, possui uma distribuição de probabilidades. Em particular, a média amostral $\overline{X}$ possui uma distribuição de probabilidades. Suponha, por exemplo, que $n = 2$ componentes são selecionados aleatoriamente, e o número de quebras durante o período de garantia é determinado para cada um. Valores possíveis para o número médio de quebras $\overline{X}$ da amostra são 0 (se $X_1 = X_2 = 0$), 0,5 (se $X_1 = 0$ e $X_2 = 1$ ou $X_1 = 1$ e $X_2 = 0$), 1, 1,5, ...). A distribuição de probabilidade de $\overline{X}$ especifica $P(\overline{X} = 0)$, $P(\overline{X} = 0{,}5)$, e assim por diante, a partir das quais outras probabilidades, como $P(1 \le \overline{X} \le 3)$ e $P(\overline{X} \ge 2{,}5)$ podem ser calculadas. De forma semelhante, se, para uma amostra de tamanho $n = 2$, os únicos valores possíveis da variância da amostra forem 0, 12,5 e 50 (que será o caso se X_1 e X_2 puderem adotar somente os valores 40, 45 e 50), então a distribuição de probabilidades de S^2 fornece $P(S^2 = 0)$, $P(S^2 = 12{,}5)$ e $P(S^2 = 50)$. A distribuição de probabilidade de uma estatística às vezes é denominada **distribuição de amostragem** para enfatizar que ela descreve como a estatística varia de valor em todas as amostras que possam ser selecionadas.

Amostras Aleatórias

A distribuição de probabilidade de qualquer estatística específica depende não somente da distribuição da população (normal, uniforme etc.) e do tamanho da amostra n, como também do método de amostragem. Considere a seleção de uma amostra de tamanho $n = 2$ de uma população, formada apenas dos três valores 1, 5 e 10, e suponha que a estatística de interesse seja a variância da amostra. Se a amostragem é feita "com reposição", então, $S^2 = 0$ será o resultado se $X_1 = X_2$. Entretanto, S^2 não pode ser igual a 0 se a amostragem for "sem reposição". Assim, $P(S^2 = 0) = 0$ para um método de amostragem, e essa probabilidade é positiva para o outro método. Nossa próxima definição descreve um método de amostragem geralmente encontrado (pelo menos aproximadamente) na prática.

DEFINIÇÃO Diz-se que as vas $X_1, X_2, \ldots, X_n$ formam uma **amostra aleatória** (simples) de tamanho n se

1. Os X_is são vas independentes.
2. Todo X_i possui a mesma distribuição de probabilidades.

As Condições 1 e 2 podem ser parafraseadas, dizendo-se que os X_is são *independente e identicamente distribuídos* (iid). Se a amostragem for com reposição ou de uma população (hipotética) infinita, as Condições 1 e 2 são satisfeitas exatamente. Tais condições serão satisfeitas aproximadamente se a amostragem for sem reposição, ainda que o tamanho da amostra n seja muito menor que o tamanho da população N. Na prática, se $n/N \le 0{,}05$ (no máximo 5% da população é coletada), podemos prosseguir como se os X_is formassem uma amostra aleatória. A virtude desse método de amostragem é que a distribuição de probabilidade de qualquer estatística pode ser obtida com mais facilidade do que qualquer outro método de amostragem.

Existem dois métodos gerais para se obter informações sobre a distribuição da amostragem da estatística. Um método envolve cálculos com base nas regras de probabilidade; o outro envolve a realização de um experimento de simulação.

Dedução da Distribuição da Amostragem de uma Estatística

As regras de probabilidade são usadas para obter a distribuição de uma estatística, contanto que seja uma função "bem simples" dos X_is e que haja relativamente poucos valores de X diferentes na população; ou então, que a distribuição da população possua uma forma "apropriada". Nossos dois próximos exemplos ilustram tais situações.

Exemplo 5.20

Um grande centro de manutenção de automóveis cobra US$ 40, US$ 45 e US$ 50 para regulagem de carros de quatro, seis e oito cilindros, respectivamente. Se 20% das regulagens são feitas nos carros de quatro cilindros, 30% nos carros de seis cilindros e 50% nos carros de oito cilindros, então a distribuição de probabilidade da receita de uma única regulagem selecionada aleatoriamente é dada por

x	40	45	50	
$p(x)$	0,2	0,3	0,5	com $\mu = 46{,}5$, $\sigma^2 = 15{,}25$ (5.2)

Suponha que, em certo dia, somente dois trabalhos de manutenção envolvam regulagem. Seja X_1 = receita da primeira regulagem e X_2 = a receita da segunda. Suponha que X_1 e X_2 sejam independentes, cada uma com a distribuição de probabilidades mostrada em (5.2) [de modo que X_1 e X_2 constituam uma amostra aleatória da distribuição (5.2)]. A Tabela 5.2 lista os pares possíveis (x_1, x_2), a probabilidade de cada um [calculada usando-se (5.2) e a hipótese de independência], e os valores resultantes de $\bar{x}$ e s^2. Para obter a distribuição de probabilidade de $\bar{X}$, a receita média amostral por regulagem, devemos considerar cada valor possível de $\bar{x}$ e calcular sua probabilidade. Por exemplo: $\bar{x} = 45$ ocorre três vezes na tabela com probabilidades 0,10, 0,09 e 0,10, assim

$$p_{\bar{X}}(45) = P(\bar{X} = 45) = 0{,}10 + 0{,}09 + 0{,}10 = 0{,}29$$

Tabela 5.2 Resultados, probabilidades e valores de $\bar{x}$ e s^2 do Exemplo 5.20

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$	s^2
40	40	0,04	40	0
40	45	0,06	42,5	12,5
40	50	0,10	45	50
45	40	0,06	42,5	12,5
45	45	0,09	45	0
45	50	0,15	47,5	12,5
50	40	0,10	45	50
50	45	0,15	47,5	12,5
50	50	0,25	50	0

De forma similar,

$$p_{S^2}(50) = P(S^2 = 50) = P(X_1 = 40, X_2 = 50 \quad \text{ou} \quad X_1 = 50, X_2 = 40)$$
$$= 0{,}10 + 0{,}10 = 0{,}20$$

As distribuições de amostragem completas de $\overline{X}$ e S^2 são exibidas em (5.3) e (5.4).

$\bar{x}$	40	42,5	45	47,5	50
$p_{\overline{X}}(\bar{x})$	0,04	0,12	0,29	0,30	0,25

(5.3)

s^2	0	12,5	50
$p_{S^2}(s^2)$	0,38	0,42	0,20

(5.4)

A Figura 5.7 ilustra o histograma probabilístico da distribuição original (5.2) e da distribuição $\overline{X}$ (5.3). A figura sugere, primeiro, que a média (valor esperado) da distribuição $\overline{X}$ seja igual à média 46,5 da distribuição original, uma vez que ambos os histogramas parecem estar centralizados no mesmo lugar.

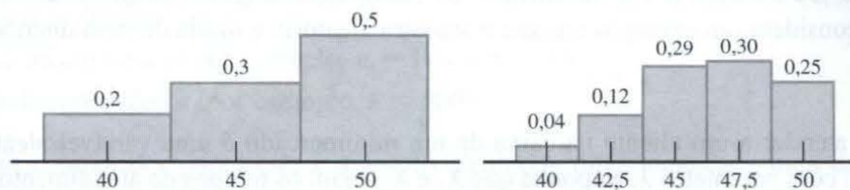


Figura 5.7 Histogramas probabilísticos da distribuição subjacente e da distribuição $\overline{X}$ do Exemplo 5.20

Pela (5.3),

$$\mu_{\overline{X}} = E(\overline{X}) = \sum \bar{x} p_{\overline{X}}(\bar{x}) = (40)(0{,}04) + \cdots + (50)(0{,}25) = 46{,}5 = \mu$$

Em segundo lugar, parece que a distribuição $\overline{X}$ possui dispersão (variabilidade) menor do que a distribuição original, uma vez que a massa de probabilidade moveu-se em direção à média. Novamente pela (5.3),

$$\begin{aligned} \sigma^2_{\overline{X}} = V(\overline{X}) &= \sum \bar{x}^2 \cdot p_{\overline{X}}(\bar{x}) - \mu^2_{\overline{X}} \\ &= (40)^2(0{,}04) + \cdots + (50)^2(0{,}25) - (46{,}5)^2 \\ &= 7{,}625 = \frac{15{,}25}{2} = \frac{\sigma^2}{2} \end{aligned}$$

A variância de $\overline{X}$ é precisamente metade da variância original (pois, $n = 2$). O valor médio de S^2 é

$$\begin{aligned} \mu_{S^2} = E(S^2) &= \sum s^2 \cdot p_{S^2}(s^2) \\ &= (0)(0{,}38) + (12{,}5)(0{,}42) + (50)(0{,}20) = 15{,}25 = \sigma^2 \end{aligned}$$

Isto é, a distribuição da amostragem $\overline{X}$ está centrada na média da população μ e a distribuição da amostragem S^2, está centrada na variância da população σ^2.

Se quatro regulagens foram feitas no dia de interesse, a receita média amostral $\overline{X}$ seria baseada em uma amostra aleatória de quatro X_is, cada um com distribuição (5.2). Eventualmente, mais cálculos produzem a fmp de $\overline{X}$ para $n = 4$ como

$\bar{x}$	40	41,25	42,5	43,75	45	46,25	47,5	48,75	50
$p_{\overline{X}}(\bar{x})$	0,0016	0,0096	0,0376	0,0936	0,1761	0,2340	0,2350	0,1500	0,0625

Pelos dados anteriores, para $n = 4$, $\mu_{\overline{X}} = 46{,}50 = \mu$ e $\sigma^2_{\overline{X}} = 3{,}8125 = \sigma^2/4$. A Figura 5.8 é um histograma de probabilidade dessa fmp.

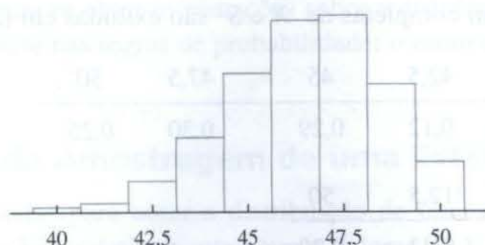


Figura 5.8 Histograma probabilístico para $\overline{X}$ com base em $n = 4$ do Exemplo 5.20 ■

O Exemplo 5.20 sugere, antes de qualquer coisa, que o cálculo de $p_{\overline{X}}(\overline{x})$ e $p_{S^2}(s^2)$ é cansativo. Se a distribuição original (5.2) permitisse mais de três valores possíveis 40, 45 e 50, então, mesmo para $n = 2$, os cálculos teriam sido mais complexos. O exemplo sugere também, no entanto, que existem algumas relações gerais entre $E(\overline{X})$, $V(\overline{X})$, $E(S^2)$, e a média μ e a variância σ^2 da distribuição original, as quais são confirmadas na próxima seção. Agora, considere um exemplo em que a amostra aleatória é tirada de uma distribuição contínua.

Exemplo 5.21

O tempo gasto para atender a um cliente na caixa de um minimercado é uma variável aleatória que tem distribuição exponencial com parâmetro λ. Suponha que X_1 e X_2 sejam os tempos de atendimento para dois clientes diferentes, assumidos como independentes. Considere o tempo total de atendimento $T_o = X_1 + X_2$ dos dois clientes também uma estatística. A fdc de T_o é para $t \geq 0$,

$$
\begin{aligned}
F_{T_o}(t) &= P(X_1 + X_2 \leq t) = \iint_{\{(x_1, x_2): x_1 + x_2 \leq t\}} f(x_1, x_2)\, dx_1\, dx_2 \\
&= \int_0^t \int_0^{t - x_1} \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2}\, dx_2\, dx_1 = \int_0^t [\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t}]\, dx_1 \\
&= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}
\end{aligned}
$$

A região de integração está ilustrada na Figura 5.9.

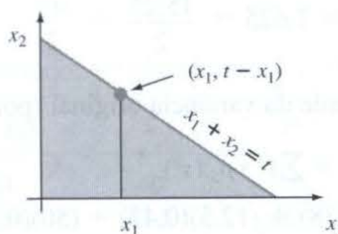


Figura 5.9 Região de integração para obter a fdc de T_o do Exemplo 5.21

A fdp de T_o é obtida diferenciando $F_{T_o}(t)$:

$$
f_{T_o}(t) = \begin{cases} \lambda^2 t e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{5.5}
$$

Essa é uma fdp gama ($\alpha = 2$ e $\beta = 1/\lambda$). A fdp de $\overline{X} = T_o/2$ é obtida da relação $\{\overline{X} \leq \overline{x}\}$ se e somente se $\{T_o \leq 2\overline{x}\}$ conforme

$$
f_{\overline{X}}(\overline{x}) = \begin{cases} 4\lambda^2 \overline{x} e^{-2\lambda \overline{x}} & \overline{x} \geq 0 \\ 0 & \overline{x} < 0 \end{cases} \tag{5.6}
$$

A média e a variância da distribuição exponencial subjacente são $\mu = 1/\lambda$ e $\sigma^2 = 1/\lambda^2$. Pelas Expressões (5.5) e (5.6), pode-se demonstrar que $E(\overline{X}) = 1/\lambda$, $V(\overline{X}) = 1/(2\lambda^2)$, $E(T_o) = 2/\lambda$ e $V(T_o) = 2/\lambda^2$. Tais resultados novamente sugerem algumas relações gerais entre médias e variâncias de $\overline{X}$, T_o e a distribuição subjacente. ■

Experimentos de Simulação

O segundo método para se obter informações sobre a distribuição de amostragem de uma estatística é fazer um experimento de simulação. Esse método geralmente é usado quando é muito difícil ou complicado calcular a derivada por meio das regras de probabilidade. Tal experimento sempre é feito virtualmente, com o auxílio de um computador. As seguintes características do experimento devem ser especificadas:

1. A estatística de interesse ($\overline{X}$, S, uma média aparada específica etc.)
2. A distribuição da população (normal com $\mu = 100$ e $\sigma = 15$, uniforme com limite inferior $A = 5$ e limite superior $B = 10$ etc.)
3. O tamanho da amostra n (por exemplo, $n = 10$ ou $n = 50$)
4. O número de repetições k (por exemplo, $k = 500$)

Então, use um computador para obter k amostras aleatórias diferentes, cada uma de tamanho n, da distribuição da população designada. Para cada amostra, calcule o valor da estatística e construa um histograma dos k valores calculados. O histograma fornece a distribuição de amostragem *aproximada* da estatística. Quanto maior o valor de k, melhor é a tendência da aproximação (a distribuição de amostragem real surge quando $k \to \infty$). Na prática, $k = 500$ ou 1000 geralmente é suficiente, se a estatística for "bem simples".

Exemplo 5.22

A distribuição da população para nosso primeiro estudo de simulação é normal com $\mu = 8{,}25$ e $\sigma = 0{,}75$, conforme ilustrado na Figura 5.10. [O artigo "Platelet Size in Myocardial Infarction" [Tamanho da Plaqueta no Infarto do Miocárdio] (*British Med. J.*, 1983, p. 449-451) sugere essa distribuição para o volume de plaquetas nos indivíduos sem histórico de problemas cardíacos graves.]

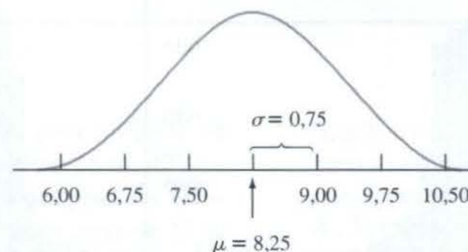


Figura 5.10 Distribuição normal, com $\mu = 8{,}25$ e $\sigma = 0{,}75$

Realmente fizemos quatro experimentos diferentes, com 500 repetições para cada um. No primeiro experimento, as 500 amostras de $n = 5$ observações cada foram geradas usando-se o MINITAB, e os tamanhos da amostra para os outros três foram $n = 10$, $n = 20$ e $n = 30$, respectivamente. A média amostral foi calculada para cada amostra e os histogramas resultantes dos valores de $\bar{x}$ são exibidos na Figura 5.11.

A primeira coisa a se observar sobre os histogramas é seu formato. Para uma aproximação razoável, cada quatro aproximações parecem com uma curva normal. A semelhança seria ainda mais impressionante se todo histograma fosse feito com base em mais de 500 valores de $\bar{x}$. Em segundo lugar, cada histograma está centrado aproximadamente em 8,25, a média da população que está servindo como amostra. Se os histogramas se baseassem em uma seqüência sem fim de valores de $\bar{x}$, seus centros seriam exatamente a média da população, 8,25.

O aspecto final a ser observado nos histogramas é sua dispersão relativa entre eles. Quanto menor o valor de n, maior a extensão com que a distribuição da amostragem se dispersa em relação ao valor médio. É por esse motivo que os histogramas de $n = 20$ e $n = 30$ baseiam-se em intervalos de classe menores que os dos dois

tamanhos menores da amostra. Para os tamanhos de amostra maiores, a maior parte dos valores de $\bar{x}$ está bem próxima de 8,25. Esse é o efeito do cálculo da média. Quando n é pequeno, um único valor de x incomum pode resultar em um valor de $\bar{x}$ longe do centro. Com um tamanho amostral maior, quaisquer valores de x incomuns quando incluídos no cálculo da média com os demais valores da amostra, ainda tendem a produzir um valor de $\bar{x}$ próximo a μ. Combinar essas percepções produz um resultado que desperta sua intuição: **$\bar{X}$ determinado com base em n grande tende a estar mais próximo de μ do que se fosse determinado com base em n pequeno.**

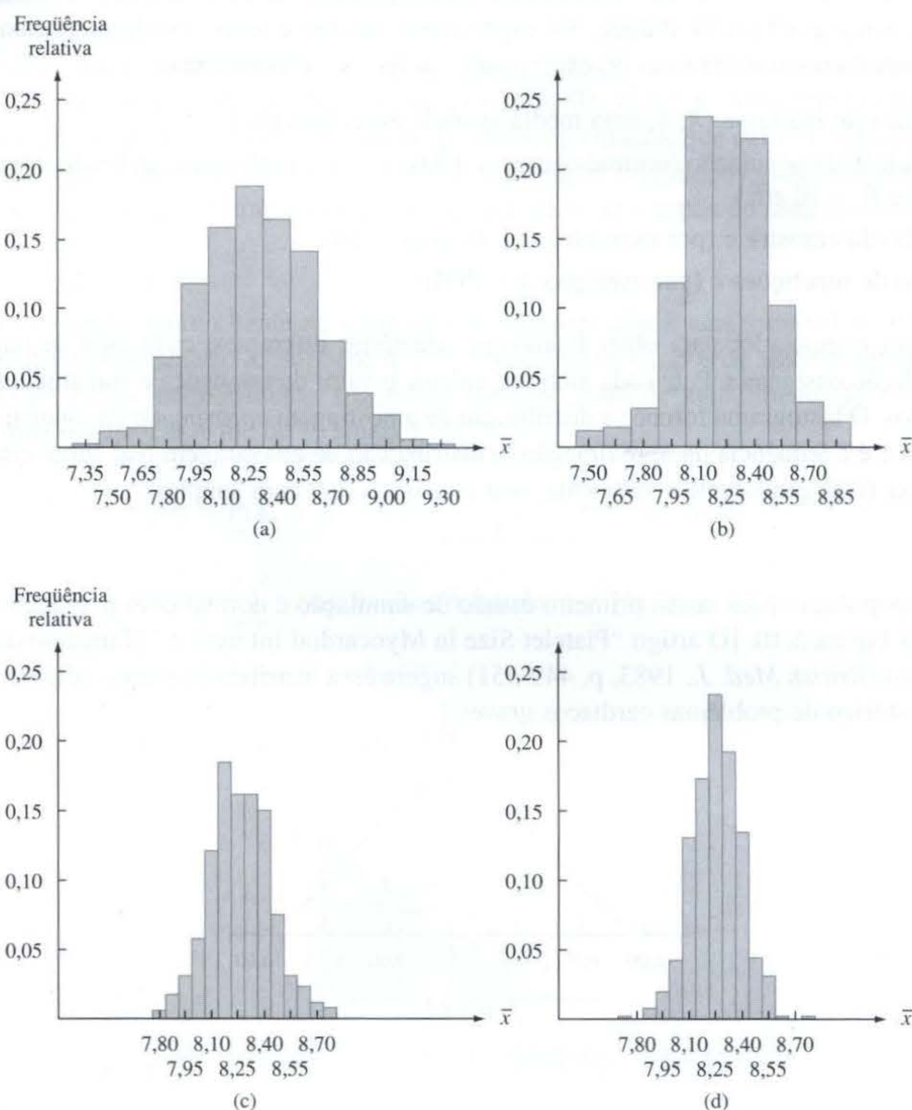


Figura 5.11 Histogramas da amostra de $\bar{x}$ com base em 500 amostras, cada uma consistindo de n observações: (a) $n = 5$; (b) $n = 10$; (c) $n = 20$; (d) $n = 30$ ■

Exemplo 5.23

Considere um experimento de simulação em que a distribuição da população seja bastante inclinada. A Figura 5.12 mais adiante mostra a curva de densidade da vida útil de determinado tipo de controle eletrônico [esta curva é realmente uma distribuição lognormal com $E(\ln(X)) = 3$ e $V(\ln(X)) = 0,4$]. Novamente, a estatística de interesse é a média amostral $\bar{X}$. O experimento utilizou 500 repetições e considerou os mesmos quatro tamanhos da amostra, como no Exemplo 5.22. Os histogramas resultantes, e um gráfico de probabilidade normal pelo MINITAB para os 500 valores de x com base em $n = 30$, são exibidos na Figura 5.13.

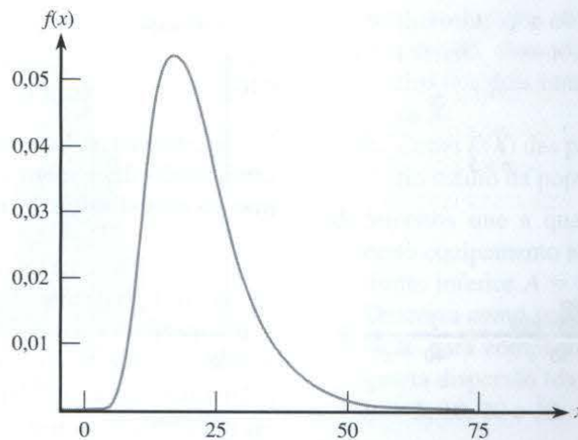


Figura 5.12 Curva de densidade para o experimento de simulação do Exemplo 5.23 [$E(X) = \mu = 21{,}7584$, $V(X) = \sigma^2 = 82{,}1449$]

Ao contrário do caso normal, esses histogramas diferem no formato. Em particular, ficam progressivamente menos inclinados à medida que o tamanho da amostra n aumenta. A média dos 500 valores de $\bar{x}$ para os quatro tamanhos diferentes da amostra está bem próxima do valor médio da distribuição da população. Se cada histograma fosse baseado em uma seqüência sem fim de valores de $\bar{x}$, não apenas 500, os quatro seriam centrados exatamente em 21,7584. Dessa forma, valores diferentes de n alteram o formato, mas não o centro da distribuição de $\bar{X}$. A comparação dos quatro histogramas na Figura 5.13 também mostra que, à medida que n aumenta, a dispersão dos histogramas diminui. Aumentar n resulta em um grau maior de concentração, em torno do valor médio da população, e faz o histograma se parecer mais com uma curva normal. O histograma da Figura 5.13(d) e o gráfico de probabilidade normal na Figura 5.13(e) fornecem evidência convincente de que o tamanho de uma amostra de $n = 30$ é suficiente para superar a inclinação da distribuição da população e dar uma distribuição de amostragem de $\bar{X}$ aproximadamente normal. ■

Exercícios | Seção 5.3 (37–45)

37. Uma determinada marca de sabão para máquina de lavar louça é vendida em três tamanhos: 25 oz, 40 oz e 65 oz. Vinte por cento de todos os compradores escolhem a caixa de 25 oz, 50% escolhem a caixa de 40 oz e os 30% restantes escolhem a caixa de 65 oz. Sejam X_1 e X_2 os tamanhos dos pacotes escolhidos por dois compradores selecionados independentemente.

a. Determine a distribuição de $\bar{X}$, calcule $E(\bar{X})$ e compare a μ.

b. Determine a distribuição da amostragem da variância da amostra S^2, calcule $E(S^2)$ e compare com σ^2.

38. Existem dois semáforos a caminho do meu trabalho. Seja X_1 o número de sinais de trânsito em que eu tenho de parar, e suponha que a distribuição de X_1 seja a seguinte:

x_1	0	1	2
$p(x_1)$	0,2	0,5	0,3

$\mu = 1{,}1$, $\sigma^2 = 0{,}49$

Seja X_2 o número de faróis de trânsito em que eu tenho de parar a caminho de casa; X_2 é independente de X_1. Assuma que X_2 possui a mesma distribuição que X_1, de modo que X_1, X_2 é uma amostra aleatória $n = 2$.

a. Seja $T_o = X_1 + X_2$ e determine a distribuição de probabilidades de T_o.

b. Calcule μ_{T_o}. Como ele está relacionado a μ, a média da população?

c. Calcule $\sigma^2_{T_o}$. Como ele está relacionado a σ^2, variância da população?

39. Sabe-se que 80% de todos os *zip drives* de marca A funcionam de forma satisfatória durante o período de garantia (são "sucessos"). Suponha que $n = 10$ *drives* sejam selecionados aleatoriamente. Seja X = número de sucessos na amostra. A estatística X/n é a proporção da amostra (fração) de sucessos. Obtenha a distribuição de amostragem dessa estatística. [*Sugestão*: um valor possível de X/n é 0,3, correspondente a $X = 3$. Qual é a probabilidade desse valor (que tipo de variável aleatória é X)?]

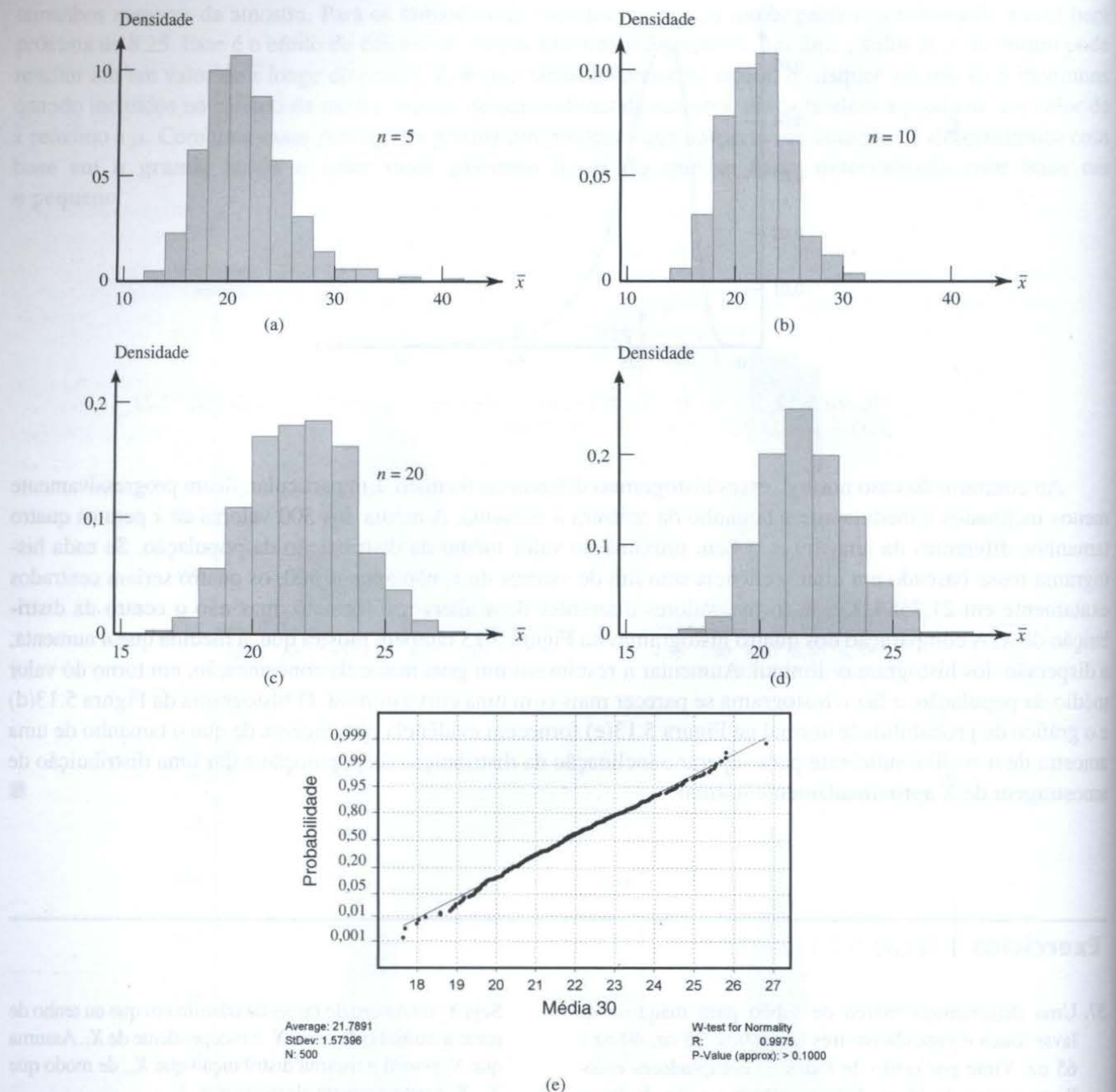


Figura 5.13 Resultados do experimento de simulação do Exemplo 5.23: (a) histograma de $\bar{x}$ para $n = 5$; (b) histograma de $\bar{x}$ para $n = 10$; (c) histograma de $\bar{x}$ para $n = 20$; (d) histograma de $\bar{x}$ para $n = 30$; (e) gráfico de probabilidade normal para $n = 30$ (do MINITAB)

40. Uma caixa contém 10 envelopes fechados numerados de 1,..., 10. Os cinco primeiros não têm dinheiro, os três seguintes têm US$ 5 cada e, em cada um dos dois últimos, há uma nota de US$ 10. Uma amostra de tamanho 3 é selecionada *com* reposição para obtermos uma amostra aleatória, e você ganha a quantia maior em qualquer um dos envelopes escolhidos. Se X_1, X_2 e X_3 representam as quantias nos envelopes selecionados, a estatística de interesse é M = o máximo de X_1, X_2 e X_3.

a. Obtenha a distribuição de probabilidades dessa estatística.

b. Descreva como você conduziria um experimento de simulação para comparar as distribuições de M dos diversos tamanhos da amostra. Como você faria para adivinhar a mudança da distribuição à medida que n aumenta?

41. Seja X o número de pacotes que estão sendo enviados por um cliente selecionado aleatoriamente em uma determinada empresa de remessas. Suponha que a distribuição de X seja como segue:

x	1	2	3	4
$p(x)$	0,4	0,3	0,2	0,1

a. Considere uma amostra aleatória de tamanho $n = 2$ (dois clientes) e seja $\overline{X}$ o número médio da amostra de pacotes enviados. Obtenha a distribuição de probabilidades de $\overline{X}$.

b. Refira-se à parte (a) e calcule $P(\overline{X} \leq 2{,}5)$.

c. Considere novamente uma amostra aleatória de tamanho $n = 2$, mas agora enfoque a estatística R = o intervalo da amostra (diferença entre os valores maior e menor na amostra). Obtenha a distribuição de R. [*Sugestão*: calcule o valor de R para cada resultado e use as probabilidades da parte (a).]

d. Se uma amostra aleatória de tamanho $n = 4$ for selecionada, qual é $P(\overline{X} \leq 1{,}5)$? (*Sugestão*: você não deve listar todos os resultados possíveis, somente aqueles para os quais $\bar{x} \leq 1{,}5$.)

42. Uma empresa mantém três escritórios em certa região, cada um com dois funcionários. As informações sobre os salários anuais (milhares de dólares) seguem abaixo:

Escritório	1	1	2	2	3	3
nº de funcionários	1	2	3	4	5	6
salário	29,7	33,6	30,2	33,6	25,8	29,7

a. Suponha que dois desses funcionários sejam selecionados aleatoriamente, dentre os seis (sem reposição). Determine a distribuição da amostragem $\overline{X}$ do salário médio da amostra.

b. Suponha que um dos três escritórios seja selecionado aleatoriamente. Sejam X_1 e X_2 os salários dos dois funcionários. Determine a distribuição de $\overline{X}$.

c. Como $E(\overline{X})$ das partes (a) e (b) é comparado ao salário médio da população μ?

43. Suponha que a quantidade de líquido despejada por certo equipamento seja uniformemente distribuída com limite inferior $A = 8$ oz e limite superior $B = 10$ oz. Descreva como você conduziria experimentos de simulação para comparar a distribuição de amostragem da quarta dispersão (da amostra) para tamanhos amostrais $n = 5$, 10, 20 e 30.

44. Faça um experimento de simulação usando um pacote estatístico ou outro software para estudar a distribuição de $\overline{X}$ quando a distribuição da população é Weibull com $\alpha = 2$ e $\beta = 5$, como no Exemplo 5.19. Considere os quatro tamanhos da amostra $n = 5$, 10, 20 e 30 e, em cada caso, utilize 500 repetições. Para qual desses tamanhos de amostra a distribuição de $\overline{X}$ parece ser aproximadamente normal?

45. Faça um experimento de simulação usando um pacote estatístico ou outro software para estudar a distribuição de $\overline{X}$ quando a distribuição da população é lognormal com $E(\ln(X)) = 3$ e $V(\ln(X)) = 1$. Considere os quatro tamanhos de amostra $n = 10$, 20, 30 e 50 e, em cada caso, utilize 500 repetições. Para qual desses tamanhos de amostra a distribuição de $\overline{X}$ parece ser aproximadamente normal?

5.4 A Distribuição da Média Amostral

A importância da média amostral $\overline{X}$ surge de seu uso para tirar conclusões sobre a média da população μ. Alguns dos procedimentos de inferência, usados com mais freqüência, baseiam-se nas propriedades da distribuição de $\overline{X}$. A apresentação inicial dessas propriedades foi feita nos cálculos e experimentos de simulação da seção anterior, onde salientamos as relações entre $E(\overline{X})$ e μ e também entre $V(\overline{X})$, σ^2 e n.

PROPOSIÇÃO Sejam $X_1, X_2, \ldots, X_n$ elementos da amostra aleatória de uma distribuição com valor médio μ e desvio padrão σ. Então,

1. $E(\overline{X}) = \mu_{\overline{X}} = \mu$
2. $V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n$ e $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

Além disso, com $T_o = X_1 + \cdots + X_n$ (o total da amostra), $E(T_o) = n\mu$, $V(T_o) = n\sigma^2$ e $\sigma_{T_o} = \sqrt{n}\sigma$.

As demonstrações desses resultados foram transferidas para a próxima seção. De acordo com o Resultado 1, a distribuição (isto é, probabilidade) da amostragem de $\overline{X}$ é centrada precisamente na média da população da qual a amostra foi selecionada. O Resultado 2 mostra que a distribuição de $\overline{X}$ se torna mais concentrada em torno de μ à medida que o tamanho da amostra n aumenta. Com diferença marcante, a distribuição de T_o se dispersa mais à medida que n aumenta. A média desloca a probabilidade para o meio, enquanto o total dispersa a probabilidade em um intervalo de valores mais amplo.

Exemplo 5.24

Em um teste de fadiga à tração de um espécime de titânio, o número esperado de ciclos para a primeira emissão acústica (usada para indicar o início da trinca) é $\mu = 28.000$, e o desvio padrão do número de ciclos é $\sigma = 5000$. Sejam $X_1, X_2, \ldots, X_{25}$ os elementos de uma amostra aleatória de tamanho 25, em que cada X_i é o número de ciclos em um espécime diferente, selecionado aleatoriamente. Então, o valor esperado do número de ciclos médio da amostra até a primeira emissão é $E(\overline{X}) = \mu = 28.000$, e o número total esperado de ciclos para os 25 espécimes é $E(T_o) = n\mu = 25(28.000) = 700.000$. Os desvios padrão de $\overline{X}$ e T_o são

$$\sigma_{\overline{X}} = \sigma/\sqrt{n} = \frac{5000}{\sqrt{25}} = 1000$$

$$\sigma_{T_o} = \sqrt{n}\sigma = \sqrt{25}(5000) = 25.000$$

Se o tamanho da amostra aumenta para $n = 100$, $E(\overline{X})$ fica inalterado, mas $\sigma_{\overline{X}} = 500$, metade de seu valor anterior (o tamanho da amostra deve ser quadruplicado para reduzir à metade o desvio padrão de $\overline{X}$). ■

Caso de Distribuição da População Normal

Ao reexaminar o experimento de simulação do Exemplo 5.22, observamos que quando a distribuição da população é normal, cada histograma de valores de $\bar{x}$ é bem aproximado pela curva normal. Segue o resultado exato.

PROPOSIÇÃO

> Sejam $X_1, X_2, \ldots, X_n$ os elementos da amostra aleatória de uma distribuição *normal* com média μ e desvio padrão σ. Então, para *qualquer n*, $\overline{X}$ é normalmente distribuído (com média μ e desvio padrão $\sigma/\sqrt{n}$), como é T_o (com média $n\mu$ e desvio padrão $\sqrt{n}\sigma$).[1]

Sabemos tudo sobre as distribuições de $\overline{X}$ e T_o quando a distribuição da população é normal. Em particular, as probabilidades, como $P(a \leq \overline{X} \leq b)$ e $P(c \leq T_o \leq d)$, podem ser obtidas simplesmente pela padronização. A Figura 5.14 ilustra a proposição.

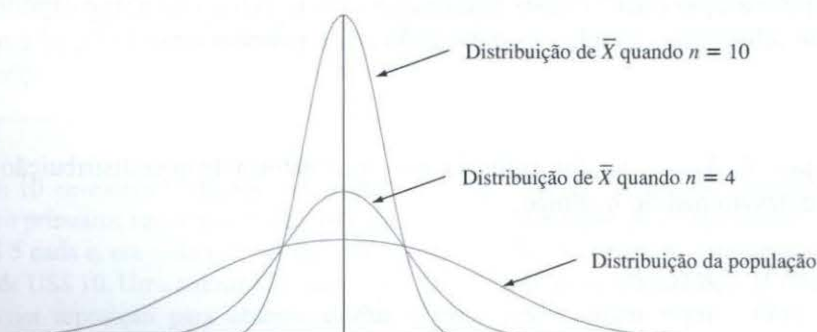


Figura 5.14 Distribuição da população normal e distribuições de amostragem de $\overline{X}$

[1] É possível demonstrar o resultado de T_o quando $n = 2$, usando o método no Exemplo 5.21, mas os detalhes são confusos. A fórmula geral normalmente é demonstrada usando-se uma ferramenta teórica denominada *função de geração de momento*. Uma das referências do capítulo pode ser consultada para obter-se mais informações.

Exemplo 5.25

O tempo que um rato de determinada subespécie, selecionado aleatoriamente leva para encontrar o caminho em um labirinto é uma va distribuída normalmente com $\mu = 1,5$ min e $\sigma = 0,35$ min. Suponha que cinco ratos sejam selecionados. Sejam $X_1,..., X_5$ seu tempo no labirinto. Assumindo que X_i é uma amostra aleatória dessa distribuição normal, qual é a probabilidade de o tempo total $T_o = X_1 + \cdots + X_5$ dos cinco estar entre 6 e 8 min? Pela proposição, T_o possui distribuição normal com $\mu_{T_o} = n\mu = 5(1,5) = 7,5$ e variância $\sigma^2_{T_o} = n\sigma^2 = 5(0,1225) = 0,6125$, assim, $\sigma_{T_o} = 0,783$. Para padronizar T_o, subtraia μ_{T_o} e divida por σ_{T_o}:

$$P(6 \leq T_o \leq 8) = P\left(\frac{6 - 7,5}{0,783} \leq Z \leq \frac{8 - 7,5}{0,783}\right)$$

$$= P(-1,92 \leq Z \leq 0,64) = \Phi(0,64) - \Phi(-1,92) = 0,7115$$

A determinação da probabilidade de o tempo médio da amostra $\overline{X}$ (uma variável distribuída normalmente) ser no máximo 2,0 min requer $\mu_{\overline{X}} = \mu = 1,5$ e $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 0,35/\sqrt{5} = 0,1565$. Então,

$$P(\overline{X} \leq 2,0) = P\left(Z \leq \frac{2,0 - 1,5}{0,1565}\right) = P(Z \leq 3,19) = \Phi(3,19) = 0,9993 \quad ■$$

Teorema do Limite Central

Quando os X_is são distribuídos normalmente, $\overline{X}$ também é para cada tamanho de amostra *n*. O experimento de simulação do Exemplo 5.23 sugere que, mesmo quando a distribuição da população é altamente não-normal, o cálculo da média produz uma distribuição em forma de sino mais acentuada do que a que está servindo como amostra. Uma hipótese razoável é que, se *n for* grande, uma curva normal adequada aproximará a distribuição real de $\overline{X}$. O enunciado formal de tal resultado é o teorema mais importante da probabilidade.

TEOREMA **Teorema do Limite Central (TLC)**

$X_1, X_2, \ldots, X_n$ formam a amostra aleatória de uma distribuição com média μ e variância σ^2. Então, se *n* é suficientemente grande, $\overline{X}$ tem aproximadamente uma distribuição normal com $\mu_{\overline{X}} = \mu$ e $\sigma^2_{\overline{X}} = \sigma^2/n$, e T_o também tem aproximadamente uma distribuição normal com $\mu_{T_o} = n\mu$, $\sigma^2_{T_o} = n\sigma^2$. Quanto maior o valor de *n*, melhor a aproximação.

A Figura 5.15 ilustra o Teorema do Limite Central. De acordo com o TLC, quando *n é* grande e queremos calcular a probabilidade, tal como $P(a \leq \overline{X} \leq b)$, precisamos somente "fingir" que $\overline{X}$ é normal, padronizá-lo e usar a tabela normal. A resposta resultante será aproximadamente correta. A resposta exata pode ser obtida somente encontrando, primeiro, a distribuição de $\overline{X}$, assim, o TLC fornece um atalho de fato impressivo. A prova do teorema envolve matemática muito avançada.

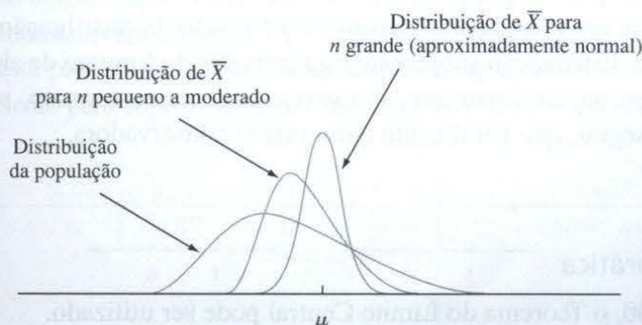


Figura 5.15 O Teorema do Limite Central ilustrado

Exemplo 5.26

Quando um lote de certo produto químico é preparado, a quantidade de uma impureza específica no lote é uma variável aleatória com valor médio de 4,0 g e desvio padrão de 1,5 g. Se 50 lotes forem preparados independentemente, qual é a probabilidade (aproximada) de a quantidade média de impureza $\bar{X}$ da amostra estar entre 3,5 e 3,8 g? De acordo com a regra prática dita em poucas palavras, $n = 50$ é grande o suficiente para que o TLC seja aplicável. $\bar{X}$ então possui aproximadamente uma distribuição normal com valor médio $\mu_{\bar{X}} = 4{,}0$ e $\sigma_{\bar{X}} = 1{,}5/\sqrt{50} = 0{,}2121$, assim,

$$P(3{,}5 \leq \bar{X} \leq 3{,}8) \approx P\left(\frac{3{,}5 - 4{,}0}{0{,}2121} \leq Z \leq \frac{3{,}8 - 4{,}0}{0{,}2121}\right)$$

$$= \Phi(-0{,}94) - \Phi(-2{,}36) = 0{,}1645 \quad ■$$

Exemplo 5.27

Uma determinada organização de consumidores normalmente reporta o número de defeitos graves de cada carro novo examinado. Suponha que o número de tais defeitos de certo modelo seja uma variável aleatória com valor médio 3,2 e desvio padrão 2,4. Dentre 100 carros selecionados aleatoriamente desse modelo, qual é a probabilidade de o número médio da amostra de defeitos graves exceder 4? Seja X_i o número de defeitos graves do i-ésimo carro na amostra aleatória. Observe que X_i é uma va discreta, mas que o TLC é aplicável, se a variável de interesse for discreta ou contínua. Além disso, apesar de o desvio padrão dessa variável não-negativa ser grande em relação ao valor médio sugerir que sua distribuição é positivamente inclinada, o tamanho grande de amostra implica que $\bar{X}$ tem a distribuição aproximadamente normal. Usando $\mu_{\bar{X}} = 3{,}2$ e $\sigma_{\bar{X}} = 0{,}24$,

$$P(\bar{X} > 4) \approx P\left(Z > \frac{4 - 3{,}2}{0{,}24}\right) = 1 - \Phi(3{,}33) = 0{,}0004 \quad ■$$

O TLC fornece compreensão clara do motivo pelo qual muitas variáveis aleatórias têm distribuições de probabilidade aproximadamente normais. Por exemplo: o erro de medição em um experimento científico pode ser visto como a soma de um número de irregularidades e erros subjacentes de pouca importância.

Embora a utilidade do TLC para inferência seja, em breve, aparente, o teor intuitivo do resultado faz muitos estudantes iniciantes sentirem dificuldade. Voltando à Figura 5.7, o histograma probabilístico à esquerda é um retrato da distribuição que está servindo como amostra. É discreta e bem inclinada, de modo que não parece ser uma distribuição normal. A distribuição de $\bar{X}$ para $n = 2$ começa a exibir certa simetria, e isso é até mais pronunciado para $n = 4$ na Figura 5.8. A Figura 5.16 contém a distribuição de probabilidades de $\bar{X}$ para $n = 8$, bem como um histograma probabilístico para essa distribuição. Com $\mu_{\bar{X}} = \mu = 46{,}5$ e $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 3{,}905/\sqrt{8} = 1{,}38$, se ajustarmos uma curva normal com tal média e desvio padrão por meio do histograma de $\bar{X}$, as áreas de retângulos do histograma probabilístico estarão razoavelmente bem aproximadas pelas áreas da curva normal, pelo menos na parte central da distribuição. O desenho de T_o é semelhante, exceto pelo fato de a escala horizontal ser muito mais dispersa, com T_o variando de 320 ($\bar{x} = 40$) a 400 ($\bar{x} = 50$).

Uma dificuldade prática na aplicação do TLC é saber quando n é suficientemente grande. O problema é que a precisão da aproximação de um n específico depende do formato da distribuição subjacente original que está servindo como amostra. Se a distribuição subjacente está próxima do formato de sino, então a aproximação será boa, mesmo para um n pequeno, ao passo que, se estiver longe desse formato, será necessário um n grande. Usaremos a regra prática a seguir, que geralmente é um pouco conservadora.

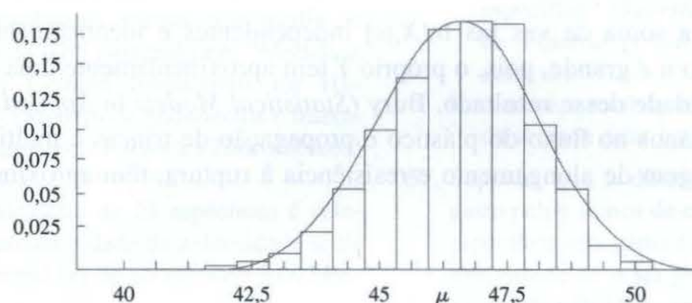
Regra prática

Se $n > 30$, o Teorema do Limite Central pode ser utilizado.

Existem distribuições de população para as quais mesmo com n 40 ou 50 não é suficiente, mas tais distribuições raramente são encontradas na prática.

$\bar{x}$	40	40,625	41,25	41,875	42,5	43,125
$p(\bar{x})$	0,0000	0,0000	0,0003	0,0012	0,0038	0,0112
$\bar{x}$	43,75	44,375	45	45,625	46,25	46,875
$p(\bar{x})$	0,0274	0,0556	0,0954	0,1378	0,1704	0,1746
$\bar{x}$	47,5	48,125	48,75	49,375	50	
$p(\bar{x})$	0,1474	0,0998	0,0519	0,0188	0,0039	

(a)



(b)

Figura 5.16 (a) Distribuição de probabilidades de $\bar{X}$ para $n = 8$; (b) Histograma probabilístico e aproximação normal à distribuição de $\bar{X}$ quando a distribuição original é como no Exemplo 5.20

Outras Aplicações do Teorema do Limite Central

O TLC pode ser usado para justificar a aproximação normal da distribuição binomial discutida no Capítulo 4. Lembre-se de que uma variável binomial X é o número de sucessos em um experimento binomial que consiste de n tentativas independentes de sucesso/falha com $p = P(S)$ para qualquer tentativa específica. Defina novas vas $X_1, X_2, \ldots, X_n$ por

$$X_i = \begin{cases} 1 & \text{se a } i\text{-ésima tentativa for um sucesso} \\ 0 & \text{se a } i\text{-ésima tentativa for um fracasso} \end{cases} \quad (i = 1, \ldots, n)$$

Em virtude de as tentativas serem independentes e $P(S)$ ser constante de tentativa para tentativa, os X_is são iid (uma amostra aleatória de uma distribuição de Bernoulli). O TLC, então, implica que, se n for suficientemente grande, tanto a soma como a média dos X_is têm aproximadamente distribuições normais. Quando os X_is são somados, adiciona-se 1 a cada S que ocorre e 0 a cada F, assim, $X_1 + \cdots + X_n = X$. A média amostral dos X_is é X/n, a proporção da amostra de sucessos. Isso é, X e X/n são aproximadamente normais quando n é grande. O tamanho necessário da amostra para essa aproximação depende do valor de p: quando p está próximo de 0,5, a distribuição de cada X_i é razoavelmente simétrica (veja a Figura 5.17 na próxima página), ao passo que a distribuição é relativamente inclinada quando p está próximo de 0 ou 1. Usar a aproximação somente se $np \geq 10$ e $n(1 - p) \geq 10$ garante que n é grande o suficiente para superar qualquer inclinação na distribuição de Bernoulli subjacente.

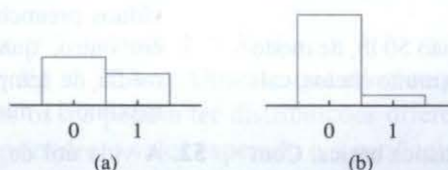


Figura 5.17 Duas distribuições de Bernoulli: (a) $p = 0,4$ (razoavelmente simétrica); (b) $p = 0,1$ (muito inclinada)

Lembre-se da Seção 4.5, em que X possui uma distribuição lognormal se $\ln(X)$ tiver distribuição normal.

PROPOSIÇÃO Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma distribuição para a qual somente valores positivos são possíveis $[P(X_i > 0) = 1]$. Então, se n for suficientemente grande, o produto $Y = X_1X_2 \cdot \cdots \cdot X_n$ terá aproximadamente uma distribuição lognormal.

Para demonstrar a proposição, observe que

$$\ln(Y) = \ln(X_1) + \ln(X_2) + \cdots + \ln(X_n)$$

Uma vez que $\ln(Y)$ é a soma de vas [as $\ln(X_i)$s] independentes e identicamente distribuídas, é aproximadamente normal quando *n é* grande, pois, o próprio Y tem aproximadamente uma distribuição lognormal. Como exemplo da aplicabilidade desse resultado, Bury *(Statistical Models in Applied Science, Wiley*, p. 590) discute que o processo de danos no fluxo do plástico e propagação de trincas é multiplicativo, de modo que as variáveis, como porcentagem de alongamento e resistência à ruptura, têm aproximadamente distribuições lognormais.

Exercícios | Seção 5.4 (46–57)

46. O diâmetro interno de um anel de pistão selecionado casualmente é uma variável aleatória com valor médio de 12 cm e desvio padrão de 0,04 cm.

a. Se $\overline{X}$ é o diâmetro médio para uma amostra aleatória de $n = 16$ anéis, onde a distribuição de amostragem de $\overline{X}$ está centrada, qual é o desvio padrão da distribuição de $\overline{X}$?

b. Responda as questões propostas na parte (a) para um tamanho de amostra de $n = 64$ anéis.

c. Para qual das duas amostras aleatórias, a da parte (a) ou a da parte (b), $\overline{X}$ é mais provável de estar dentro de 0,01 cm de 12 cm? Explique seu raciocínio.

47. Consulte o Exercício 46. Suponha que a distribuição do diâmetro seja normal.

a. Calcule $P(11{,}99 \le \overline{X} \le 12{,}01)$ quando $n = 16$.

b. Qual é a probabilidade de o diâmetro médio da amostra exceder 12,01 quando $n = 25$?

48. Sejam $X_1, X_2, \ldots, X_{100}$ os pesos líquidos reais de 100 sacos de fertilizantes de 50 lb selecionados aleatoriamente.

a. Se o peso esperado de cada saco for 50 e a variância 1, calcule $P(49{,}75 \le \overline{X} \le 50{,}25)$ (aproximadamente), usando o TLC.

b. Se o peso esperado for 49,8 lb, e não 50 lb, de modo que na média os sacos não estejam muito cheios, calcule $P(49{,}75 \le \overline{X} \le 50{,}25)$.

49. Há 40 alunos em uma classe de estatística básica. Com base nos anos de experiência, o instrutor sabe que o tempo necessário para corrigir o primeiro questionário, escolhido aleatoriamente, é uma variável aleatória com valor esperado de 6 min e desvio padrão de 6 min.

a. Se os tempos de correção forem independentes e o instrutor começar a corrigir às 18h50 sem parar, qual é a probabilidade (aproximada) de terminar as correções antes de o noticiário das 23h, na televisão, começar?

b. Se as notícias de esportes começam às 23h10, qual é a probabilidade de ele perder parte das notícias, se esperar até que a correção seja concluída antes de ligar a televisão?

50. A resistência à ruptura de um rebite possui valor médio de 10.000 psi e desvio padrão de 500 psi.

a. Qual é a probabilidade de a resistência à ruptura média de uma amostra aleatória de 40 rebites estar entre 9.900 e 10.200?

b. Se o tamanho da amostra fosse 15, e não 40, a probabilidade pedida na parte (a) poderia ser calculada com as informações dadas?

51. O tempo que um candidato a uma hipoteca, selecionado aleatoriamente, gasta para preencher um formulário tem distribuição normal com valor médio de 10 min e desvio padrão de 2 min. Se cinco indivíduos preenchem um formulário em um dia, e seis, em outro, qual é a probabilidade de a quantidade média de tempo da amostra em cada dia ser de no máximo 11 min?

52. A vida útil de certo tipo de bateria é distribuída normalmente com valor médio de 10 horas e desvio padrão de 1 hora. Há quatro baterias em um pacote. Que valor

de vida útil faz com que a vida total de todas as baterias em um pacote exceda tal valor em apenas 5% de todos os pacotes?

53. É sabido que a dureza de Rockwell de certo tipo de pinos tem valor médio de 50 e desvio padrão de 1,2.
   a. Se a distribuição for normal, qual é a probabilidade de a dureza média de uma amostra aleatória de 9 (nove) pinos ser no mínimo 51?
   b. Qual é a probabilidade (aproximada) de a dureza média de uma amostra aleatória de 40 pinos ser de no mínimo 51?

54. Suponha que a densidade sedimentar (g/cm) de um espécime selecionado aleatoriamente de uma determinada região tenha distribuição normal com média de 2,65 e desvio padrão de 0,85 (sugerido em "Modeling Sediment and Water Column Interactions for Hydrophobic Pollutants" (Modelo de Sedimentos e Interações da Coluna de Água para Poluentes Hidrofóbicos) *Water Research,* 1984, p. 1169-1174).
   a. Se uma amostra aleatória de 25 espécimes é selecionada, qual é a probabilidade de a densidade sedimentar média amostral ser de no máximo 3,00? Entre 2,65 e 3,00?
   b. Qual deve ser o tamanho de uma amostra para garantir que a primeira probabilidade da parte (a) seja no mínimo 0,99?

55. O primeiro trabalho de uma classe de cálculo estatístico envolve a execução de um programa curto. Se a experiência anterior indica que 40% de todos os alunos não cometerão erros de programação, calcule a probabilidade (aproximada) de que em uma classe de 50 alunos:
   a. pelo menos 25 não cometerão erros. (*Sugestão*: aproximação da normal para a binomial.)
   b. entre 15 e 25 (inclusive) não cometerão erros.

56. O número de multas por estacionamento irregular emitidas em uma determinada cidade em um dia de semana qualquer possui distribuição de Poisson com parâmetro $\lambda = 50$. Qual é a probabilidade aproximada de
   a. entre 35 e 70 multas serem dadas em um dia específico? (*Sugestão*: quando λ é grande, uma va de Poisson possui aproximadamente distribuição normal.)
   b. o número total de multas dadas durante os cinco dias da semana estar entre 225 e 275?

57. Suponha que a distribuição do tempo X (em horas) gasto pelos alunos de certa universidade em um projeto específico seja gama com parâmetros $\alpha = 50$ e $\beta = 2$. Em virtude de α ser grande, pode-se demonstrar que X possui distribuição aproximadamente normal. Use esse fato para calcular a probabilidade de um aluno selecionado aleatoriamente gastar no máximo 125 horas no projeto.

5.5 Distribuição de uma Combinação Linear

A $\bar{X}$ média amostral e o T_o total da amostra são casos especiais de um tipo de variável aleatória que surge com bastante freqüência em aplicações estatísticas.

DEFINIÇÃO

Dado um conjunto de n variáveis aleatórias $X_1, \ldots, X_n$ e n constantes numéricas $a_1, \ldots, a_n$, a va

$$Y = a_1X_1 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i \tag{5.7}$$

é denominada **combinação linear** dos X_i's.

Assumindo $a_1 = a_2 = \cdots = a_n = 1$ fornece $Y = X_1 + \cdots + X_n = T_o$, e $a_1 = a_2 = \cdots = a_n = \frac{1}{n}$ produz $Y = \frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}T_o = \bar{X}$. Observe que não é requerido que os X_is sejam independentes ou identicamente distribuídos. Todos os X_is podem ter distribuições diferentes e, dessa forma, valores médios e variâncias distintas. Consideramos primeiro o valor esperado e a variância de uma combinação linear.

PROPOSIÇÃO Suponha que $X_1, X_2 \ldots X_n$ tenham valores médios $\mu_1, \ldots, \mu_n$, respectivamente, e variâncias $\sigma_1^2, \ldots, \sigma_n^2$, respectivamente.

1. Se os X_is são ou não independentes,

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n) = a_1\mu_1 + \cdots + a_n\mu_n \quad (5.8)$$

2. Se $X_1, \ldots, X_n$ são independentes,

$$V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n) = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2 \quad (5.9)$$

e

$$\sigma_{a_1X_1+\cdots+a_nX_n} = \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2} \quad (5.10)$$

3. Para qualquer $X_1, \ldots, X_n$,

$$V(a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j\,\text{Cov}(X_i, X_j) \quad (5.11)$$

As demonstrações estão esquematicamente no final da seção. Uma paráfrase de (5.8) é que o valor desejado de uma combinação linear é a mesma combinação linear dos valores esperados - por exemplo: $E(2X_1 + 5X_2) = 2\mu_1 + 5\mu_2$. O resultado (5.9) da Proposição 2 é um caso especial de (5.11) da Proposição 3; quando os X_is são independentes, $\text{Cov}(X_i, X_j) = 0$ para $i \neq j$ e $= V(X_i)$ para $i = j$ (essa simplificação realmente ocorre quando os X_is são não-correlacionados, uma condição mais fraca que a independência). Generalizando para o caso de uma amostra aleatória (X_is iid) com $a_i = 1/n$ para cada i fornece $E(\overline{X}) = \mu$ e $V(\overline{X}) = \sigma^2/n$, conforme discutido na Seção 5.4. Aplica-se um comentário semelhante às regras de T_o.

Exemplo 5.28

Um posto de gasolina vende três tipos do produto: gasolina comum, gasolina aditivada e gasolina super. Elas custam US$ 1,20, US$ 1,35 e US$ 1,50 por galão, respectivamente. Sejam X_1, X_2, e X_3 os valores desses galões comprados em um dia específico. Suponha que os X_is sejam independentes com $\mu_1 = 1000$, $\mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$, e $\sigma_3 = 50$. A receita das vendas é $Y = 1{,}2X_1 + 1{,}35X_2 + 1{,}5X_3$, e

$$E(Y) = 1{,}2\mu_1 + 1{,}35\mu_2 + 1{,}5\mu_3 = \$\ 2325$$

$$V(Y) = (1{,}2)^2\sigma_1^2 + (1{,}35)^2\sigma_2^2 + (1{,}5)^2\sigma_3^2 = 31.689$$

$$\sigma_Y = \sqrt{31.689} = \$178{,}01$$

■

Diferença Entre Duas Variáveis Aleatórias

Um caso especial importante de combinação linear resulta de $n = 2$, $a_1 = 1$, e $a_2 = -1$:

$$Y = a_1X_1 + a_2X_2 = X_1 - X_2$$

Temos, então, a seguinte conclusão da proposição.

COROLÁRIO $E(X_1 - X_2) = E(X_1) - E(X_2)$ e, se X_1 e X_2 são independentes, $V(X_1 - X_2) = V(X_1) + V(X_2)$.

O valor esperado de uma diferença é a diferença entre os dois valores esperados, mas a variância da diferença entre duas variáveis independentes é a *soma*, *não* a diferença, das duas variâncias. Há tanta variabilidade em $X_1 - X_2$ como em $X_1 + X_2$ [escrever $X_1 - X_2 = X_1 + (-1)X_2$, $(-1)X_2$ varia tanto quanto o próprio X_2].

Exemplo 5.29

Um fabricante de automóveis equipa um determinado modelo com um motor de seis cilindros ou de quatro cilindros. Sejam por X_1 e X_2 o consumo de combustível de carros com seis e quatro cilindros selecionados de modo independente e aleatório, respectivamente. Com $\mu_1 = 22$, $\mu_2 = 26$, $\sigma_1 = 1{,}2$, e $\sigma_2 = 1{,}5$,

$$E(X_1 - X_2) = \mu_1 - \mu_2 = 22 - 26 = -4$$

$$V(X_1 - X_2) = \sigma_1^2 + \sigma_2^2 = (1{,}2)^2 + (1{,}5)^2 = 3{,}69$$

$$\sigma_{X_1 - X_2} = \sqrt{3{,}69} = 1{,}92$$

Se fizermos nova indicação de que X_1 se refere ao carro de quatro cilindros, então $E(X_1 - X_2) = 4$, mas a variância da diferença ainda será 3,69. ■

Caso de Variáveis Aleatórias Normais

Quando os X_is formam uma amostra aleatória de uma distribuição normal, $\overline{X}$ e T_o são ambos distribuídos normalmente. Eis um resultado mais geral sobre combinações lineares.

PROPOSIÇÃO Se $X_1, X_2, \ldots, X_n$ são vas independentes, distribuídas normalmente (com médias e/ou variâncias possivelmente diferentes), então qualquer combinação linear dos X_is também tem distribuição normal. Em particular, a própria diferença $X_1 - X_2$ entre duas variáveis independentes e normalmente distribuídas é normalmente distribuída.

Exemplo 5.30 (Continuação do Exemplo 5.28)

A receita total da venda dos três tipos de gasolina em um dia específico foi $Y = 1{,}2X_1 + 1{,}35X_2 + 1{,}5X_3$, e calculamos $\mu_Y = 2325$ e (assumindo a independência) $\sigma_Y = 178{,}01$. Se os X_is são distribuídos normalmente, a probabilidade de a receita exceder 2500 é

$$P(Y > 2500) = P\left(Z > \frac{2500 - 2325}{178{,}01}\right)$$

$$= P(Z > 0{,}98) = 1 - \Phi(0{,}98) = 0{,}1635$$ ■

O TLC também pode ser generalizado de modo que se aplique a certas combinações lineares. *Grosso modo*, se n for grande e nenhum termo individual igualmente provável de contribuir muito para o valor total, então Y possui aproximadamente uma distribuição normal.

Provas para o caso de $n = 2$

Para o resultado dos valores esperados, suponha que X_1 e X_2 sejam contínuas com fdp conjunta $f(x_1, x_2)$. Então

$$E(a_1X_1 + a_2X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (a_1x_1 + a_2x_2)f(x_1, x_2)\,dx_1\,dx_2$$

$$= a_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2)\,dx_2\,dx_1$$

$$+ a_2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1, x_2)\,dx_1\,dx_2$$

$$= a_1\int_{-\infty}^{\infty} x_1 f_{X_1}(x_1)\,dx_1 + a_2\int_{-\infty}^{\infty} x_2 f_{X_2}(x_2)\,dx_2$$

$$= a_1E(X_1) + a_2E(X_2)$$

A soma substitui a integração no caso discreto. O argumento para o resultado da variância não requer que seja especificado se a variável é discreta ou contínua. Lembrando que $V(Y) = E[(Y - \mu_Y)^2]$,

$$V(a_1X_1 + a_2X_2) = E\{[a_1X_1 + a_2X_2 - (a_1\mu_1 + a_2\mu_2)]^2\}$$
$$= E\{a_1^2(X_1 - \mu_1)^2 + a_2^2(X_2 - \mu_2)^2 + 2a_1a_2(X_1 - \mu_1)(X_2 - \mu_2)\}$$

A expressão dentro das chaves é uma combinação linear das variáveis $Y_1 = (X_1 - \mu_1)^2$, $Y_2 = (X_2 - \mu_2)^2$, e $Y_3 = (X_1 - \mu_1)(X_2 - \mu_2)$, assim, fazer a operação de E para os três termos resulta em $a_1^2V(X_1) + a_2^2V(X_2) + 2a_1a_2$ Cov(X_1, X_2), como requerido. ■

Exercícios Seção 5.5 (58–74)

58. Uma empresa de remessa trabalha com contêineres de três tamanhos diferentes: (1) 27 pé³ (3 × 3 × 3), (2) 125 pé³ e (3) 512 pé³. Seja X_i ($i = 1, 2, 3$) o número de contêineres de tipo i enviados durante uma determinada semana. Com $\mu_i = E(X_i)$ e $\sigma_i^2 = V(X_i)$, suponha que os valores médios e os desvios padrão sejam como segue:

$\mu_1 = 200$ $\mu_2 = 250$ $\mu_3 = 100$
$\sigma_1 = 10$ $\sigma_2 = 12$ $\sigma_3 = 8$

a. Assumindo que X_1, X_2, X_3 sejam independentes, calcule o valor esperado e a variância do volume total enviado. [*Sugestão*: volume $= 27X_1 + 125X_2 + 512X_3$.]

b. Seus cálculos estariam necessariamente corretos se os X_is não fossem independentes? Explique.

59. Sejam X_1, X_2, e X_3 os tempos necessários para realizar três reparos sucessivos em determinada oficina. Suponha que sejam vas normais independentes com valores esperados μ_1, μ_2, e μ_3 e variâncias σ_1^2, σ_2^2, e σ_3^2, respectivamente.

a. Se $\mu = \mu_2 = \mu_3 = 60$ e $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 15$, calcule $P(X_1 + X_2 + X_3 \leq 200)$. Qual é $P(150 \leq X_1 + X_2 + X_3 \leq 200)$?

b. Usando os μ_is e σ_is dados na parte (a), calcule $P(55 \leq \bar{X})$ e $P(58 \leq \bar{X} \leq 62)$.

c. Usando os μ_is e σ_is dados na parte (a), calcule $P(-10 \leq X_1 - 0{,}5X_2 - 0{,}5X_3 \leq 5)$.

d. Se $\mu_1 = 40$, $\mu_2 = 50$, $\mu_3 = 60$, $\sigma_1^2 = 10$, $\sigma_2^2 = 12$ e $\sigma_3^2 = 14$, calcule $P(X_1 + X_2 + X_3 \leq 160)$ e $P(X_1 + X_2 \geq 2X_3)$.

60. Cinco automóveis do mesmo tipo farão uma viagem de 300 milhas. Os dois primeiros usarão uma marca de gasolina econômica, e os outros três, uma marca renomada. Sejam X_1, X_2, X_3, X_4, e X_5 os consumos de combustível observados (milhas/galão) dos cinco carros. Suponha que essas variáveis sejam independentes e normalmente distribuídas com $\mu_1 = \mu_2 = 20$, $\mu_3 = \mu_4 = \mu_5 = 21$ e $\sigma^2 = 4$ para a marca econômica e 3,5 para a marca renomada. Defina a va Y como

$$Y = \frac{X_1 + X_2}{2} - \frac{X_3 + X_4 + X_5}{3}$$

de modo que Y seja a medida da diferença do consumo entre a gasolina econômica e a renomada. Calcule $P(0 \leq Y)$ e $P(-1 \leq Y \leq 1)$. [*Sugestão*: $Y = a_1X_1 + \cdots + a_5X_5$, com $a_1 = \frac{1}{2}, \ldots, a_5 = -\frac{1}{3}$.]

61. O Exercício 26 introduziu as variáveis aleatórias X e Y, número de carros e de ônibus, respectivamente, transportados por uma balsa em uma única viagem. A fmp conjunta de X e Y é dada na tabela do Exercício 7. Demonstra-se com facilidade que X e Y são independentes.

a. Calcule o valor esperado, a variância e o desvio padrão do número total de veículos em uma única viagem.

b. Se cada carro pagar um pedágio de US$ 3 e cada ônibus de US$ 10, calcule o valor esperado, a variância e o desvio padrão da receita resultante de uma única viagem.

62. A fabricação de um componente requer três operações diferentes de usinagem. O tempo de usinagem de cada operação possui distribuição normal e os três tempos são independentes um do outro. Os valores médios são 15, 30 e 20 min, respectivamente, e os desvios padrão são 1, 2 e 1,5 min, respectivamente. Qual é a probabilidade de a usinagem levar no máximo 1 hora para produzir um componente selecionado aleatoriamente?

63. Consulte o Exercício 3.

a. Calcule a covariância entre X_1 = o número de clientes na caixa expressa e X_2 = o número de clientes na caixa superexpressa.

b. Calcule $V(X_1 + X_2)$. Como este valor se compara a $V(X_1) + V(X_2)$?

64. Suponha que seu tempo de espera por um ônibus, de manhã, seja uniformemente distribuído em [0,8], enquanto à noite, é uniformemente distribuído em [0, 10] independente do tempo de espera pela manhã.

a. Se você tomar o ônibus toda manhã e toda noite por uma semana, qual é seu tempo de espera total? (*Sugestão*: Defina as vas $X_1, \ldots, X_{10}$ e use uma regra de valor esperado.)

b. Qual é a variância do seu tempo de espera total?

c. Quais são o valor esperado e a variância da diferença entre os tempos de espera pela manhã e à noite em um dia?

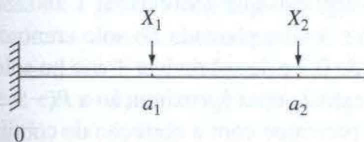
d. Quais são o valor esperado e a variância da diferença entre o tempo de espera total pela manhã e o tempo de espera total à noite, durante uma semana específica?

65. Suponha que, quando o pH de certo composto químico for 5,00, o pH medido por um aluno iniciante de química selecionado ao acaso é uma variável aleatória com média 5,00 e desvio padrão 0,2. Um grande lote do composto é subdividido e uma amostra foi dada a cada aluno do laboratório da manhã e do laboratório da tarde. Seja $\overline{X}$ = pH médio determinado pelos alunos da manhã e $\overline{Y}$ = o pH médio determinado pelos alunos da tarde.

a. Se o pH for uma variável normal e houver 25 alunos em cada laboratório, calcule $P(-0{,}1 \leq \overline{X} - \overline{Y} \leq 0{,}1)$. (*Sugestão*: $\overline{X} - \overline{Y}$ é uma combinação linear de variáveis normais, portanto, é normalmente distribuída. Calcule $\mu_{\overline{X}-\overline{Y}}$ e $\sigma_{\overline{X}-\overline{Y}}$.)

b. Se há 36 alunos em cada laboratório, mas as determinações do pH não são consideradas normais, calcule (aproximadamente) $P(-0{,}1 \leq \overline{X} - \overline{Y} \leq 0{,}1)$.

66. Se duas cargas são aplicadas a uma viga em balanço, conforme mostra o desenho a seguir, o momento fletor em 0 devido às cargas é $a_1X_1 + a_2X_2$.



a. Suponha que X_1 e X_2 sejam vas independentes com médias de 2 e 4 kips / 2000 e 4000 lb, respectivamente, e desvios padrão de 0,5 e 1,0 kip / 500 e 100 lb, respectivamente. Se a_1 = 5 ft / 1,52 m e a_2 = 10 ft / 3,04 m, qual é o momento fletor esperado e qual é o seu desvio padrão?

b. Se X_1 e X_2 são normalmente distribuídos, qual é a probabilidade de o momento fletor exceder 75 kip-ft?

c. Suponha que as posições das duas cargas sejam variáveis aleatórias. Representando-as por A_1 e A_2, assuma que essas variáveis tenham médias de 5 ft e 10 ft, respectivamente, que cada uma tenha desvio padrão de 0,5 e que todos os A_is e X_is sejam independentes uns dos outros. Qual é o momento esperado, agora?

d. Para a situação da parte (c), qual é a variância do momento fletor?

e. Se a situação for a descrita na parte (a), exceto pelo fato de que $\text{Corr}(X_1, X_2) = 0{,}5$ (de modo que as duas cargas não sejam independentes), qual é a variância do momento fletor?

67. Um pedaço de cano de PVC será colocado dentro de outro pedaço. O comprimento do primeiro pedaço é distribuído normalmente com valor médio de 20 pol e desvio padrão de 0,5 pol. O comprimento do segundo pedaço é uma va normal com média e desvio padrão de 15 pol e 0,4 pol, respectivamente. O valor da superposição é distribuído normalmente com valor médio de 1 pol e desvio padrão de 0,1 pol. Assumindo que os comprimentos e o valor da superposição sejam independentes um do outro, qual é a probabilidade de o comprimento total após a inserção estar entre 34,5 pol e 35 pol?

68. Dois aviões estão voando na mesma direção em corredores paralelos adjacentes. No tempo $t = 0$, o primeiro avião está 10 km à frente do segundo. Suponha que a velocidade do primeiro (km/h) seja distribuída normalmente com média de 520 e desvio padrão de 10, e que a velocidade do segundo avião também seja normalmente distribuída com média e desvio padrão de 500 e 10, respectivamente.

a. Qual é a probabilidade de, após 2 h de vôo, o segundo avião não ter alcançado o primeiro?

b. Determine a probabilidade de os aviões estarem separados por no máximo 10 km após 2 h.

69. Três estradas diferentes se encontram na entrada de uma determinada rodovia. Suponha que durante um período de tempo estabelecido, o número de carros que vêm de cada estrada para a rodovia seja uma variável aleatória, com valor esperado e desvio padrão dados na tabela.

	Estrada 1	Estrada 2	Estrada 3
Valor esperado	800	1000	600
Desvio padrão	16	25	18

a. Qual é o número total esperado de carros que entram na rodovia nesse ponto durante o período? (*Sugestão*: represente por X_i = o número da estrada *i*.)

b. Qual é a variância do número total de carros que entram? Você fez alguma hipótese sobre a relação entre os números de carros nas diferentes estradas?

c. Com X_i representando o número de carros que entram da estrada *i* durante o período, suponha que $\text{Cov}(X_1, X_2) = 80$, $\text{Cov}(X_1, X_3) = 90$, e $\text{Cov}(X_2, X_3) = 100$ (de modo que os três fluxos de trânsito não sejam independentes). Calcule o número total esperado de carros que entram e o desvio padrão do total.

70. Suponha que tomemos uma amostra aleatória de tamanho *n* de uma distribuição contínua que possui mediana 0, de modo que a probabilidade de qualquer observação ser positiva seja de 0,5. Agora, desprezamos os sinais das observações, classificamo-nas da menor para a maior em valor absoluto e, então, representamos por W = soma das classificações das observações que têm sinal positivo. Por exemplo: se as observações forem −0,3, +0,7, +2,1, e −2,5, então as classificações das observações positivas serão 2 e 3, assim, $W = 5$. No Capítulo 15, *W* será denominada *estatística de postos de Wilcoxon*. *W* pode ser representada como segue:

$$W = 1 \cdot Y_1 + 2 \cdot Y_2 + 3 \cdot Y_3 + \cdots + n \cdot Y_n = \sum_{i=1}^{n} i \cdot Y_i$$

onde os Y_is são vas de Bernoulli independentes, cada uma com $p = 0{,}5$ ($Y_i = 1$ corresponde à observação com classificação *i* positiva). Calcule os dados a seguir:

a. $E(Y_i)$ e, então, $E(W)$, usando a equação para *W* [*Sugestão*: a soma dos *n* primeiros inteiros positivos *é* $n(n + 1)/2$.]

b. $V(Y_i)$ e, então $V(W)$ [*Sugestão*: a soma dos quadrados dos n primeiros inteiros positivos é $n(n + 1)(2n + 1)/6$.]

71. No Exercício 66, o peso próprio da viga contribui para o momento fletor. Assuma que a viga seja de espessura e densidade uniformes, de modo que a carga resultante seja distribuída uniformemente. Se o peso da viga for aleatório, a carga resultante do peso também será; represente essa carga por W (kip-ft).

a. Se a viga tem 12 pés de comprimento, W tem média de 1,5 e desvio padrão de 0,25 e as cargas fixas são as descritas na parte (a) do Exercício 66, quais são o valor esperado e a variância do momento fletor? (*Sugestão*: se a carga, devido à viga, fosse w kip-pé, a contribuição para o momento fletor seria $w \int_0^{12} x\,dx$.)

b. Se as três variáveis (X_1, X_2, e W) são normalmente distribuídas, qual é a probabilidade de o momento fletor ser de no máximo 200 kip-ft?

72. Sou responsável por três tarefas no prédio administrativo. Seja X_i = o tempo que leva a i-ésima tarefa (i = 1, 2, 3), e X_4 = o tempo total, em minutos, que eu perco, andando pelo prédio e entre cada tarefa. Suponha que os X_is sejam independentes, normalmente distribuídos, com as seguintes médias e desvios padrão: $\mu_1 = 15$, $\sigma_1 = 4$, $\mu_2 = 5$, $\sigma_2 = 1$, $\mu_3 = 8$, $\sigma_3 = 2$, $\mu_4 = 12$, $\sigma_4 = 3$. Pretendo sair do escritório exatamente às 10 h e deixar um bilhete na porta, dizendo: "Retornarei às t h". A que horas t devo escrever, se quiser que a probabilidade da minha chegada depois de t seja 0,01?

73. Suponha que a resistência de tração esperada do aço tipo A seja 105 ksi e que o desvio padrão da resistência de tração seja 8 ksi. Para o aço tipo B, suponha que a resistência de tração esperada e o desvio padrão da resistência de tração sejam 100 ksi e 6 ksi, respectivamente. Seja $\overline{X}$ = a resistência de tração média de uma amostra aleatória de 40 espécimes tipo A, e por $\overline{Y}$ = a resistência de tração média de uma amostra aleatória de 35 espécimes tipo B.

a. Qual é a distribuição aproximada de $\overline{X}$? De $\overline{Y}$?

b. Qual é a distribuição aproximada de $\overline{X} - \overline{Y}$? Justifique sua resposta.

c. Calcule (aproximadamente) $P(-1 \le \overline{X} - \overline{Y} \le 1)$.

d. Calcule $P(\overline{X} - \overline{Y} \ge 10)$. Se realmente observou $\overline{X} - \overline{Y} \ge 10$, você duvidaria que $\mu_1 - \mu_2 = 5$?

74. Em uma área que possui solo arenoso, foram plantadas 50 árvores pequenas de um determinado tipo e outras 50 foram plantadas em uma área de solo argiloso. Sejam X = número de árvores plantadas no solo arenoso que sobrevivem 1 ano, e Y = número de árvores plantadas no solo argiloso que sobrevivem 1 ano. Se a probabilidade de a árvore plantada no solo arenoso sobreviver 1 ano for de 0,7 e de sobreviver 1 ano no solo argiloso for de 0,6, calcule uma aproximação a $P(-5 \le X - Y \le 5)$ (não se preocupe com a correção de continuidade).

Exercícios Suplementares (75–94)

75. Um restaurante serve três pratos ao preço fixo de US$ 12, US$ 15 e US$ 20. Para um casal selecionado aleatoriamente que janta nesse restaurante, sejam X = custo do jantar do homem e Y = custo do jantar da mulher. A fmp conjunta de X e Y é dada na tabela a seguir:

	$p(x, y)$	y = 12	15	20
x	12	0,05	0,05	0,10
	15	0,05	0,10	0,35
	20	0	0,20	0,10

a. Calcule as fmps marginais de X e Y.

b. Qual é a probabilidade de o jantar do homem e da mulher custar no máximo US$ 15 cada?

c. X e Y são independentes? Justifique sua resposta.

d. Qual é o custo total esperado do jantar para as duas pessoas?

e. Suponha que, quando um casal abre biscoitos da sorte no final da refeição, encontra a mensagem "Vocês receberão de volta a diferença entre o custo da refeição mais cara e o da mais barata que escolheram". Quanto o restaurante espera devolver?

76. Na estimativa de custo, o preço total de um projeto é a soma dos gastos das tarefas componentes. Cada custo é uma variável aleatória com uma distribuição de probabilidade. É costume obter informações sobre a distribuição do custo total, somando-se características das distribuições de custo do componente individual; isso é chamado procedimento de ***roll-up***. Por exemplo: $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$, assim, o procedimento de *roll-up* é válido para o custo médio. Suponha que existam duas tarefas componentes e que X_1 e X_2 sejam variáveis aleatórias independentes e normalmente distribuídas. O procedimento de *roll-up* é válido para o 75º percentil? Isto é, o 75º percentil da distribuição de $X_1 + X_2$ é o mesmo que a soma dos 75º percentis das duas distribuições individuais? Se não, qual é a relação entre o percentil da soma e a soma dos percentis? Para quais percentis o procedimento de *roll-up* é válido, nesse caso?

77. Uma loja de comida saudável possui duas marcas diferentes de um certo tipo de grão. Sejam X = quantidade (lb) da marca A em estoque e Y = quantidade da marca B em estoque. Suponha que a fdp conjunta de X e Y seja

$$f(x, y) = \begin{cases} kxy & x \ge 0,\ y \ge 0,\ 20 \le x + y \le 30 \\ 0 & \text{caso contrário} \end{cases}$$

a. Desenhe a região de densidade positiva e determine o valor de k.

b. X e Y são independentes? Responda, encontrando primeiro a fdp marginal de cada variável.

c. Calcule $P(X + Y \leq 25)$.

d. Qual é a quantidade total esperada desse grão em estoque?

e. Calcule $\text{Cov}(X, Y)$ e $\text{Corr}(X, Y)$.

f. Qual é a variância da quantidade total de grão em estoque?

78. Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias que representam n propostas independentes de um item que está à venda. Suponha que cada X_i seja uniformemente distribuído no intervalo [100, 200]. Se o vendedor vende pela oferta maior, quanto pode esperar ganhar com a venda? [*Sugestão*: seja $Y = \max(X_1, X_2, \ldots, X_n)$. Encontre primeiro $F_Y(y)$, observando que $Y \leq y$ se e somente se cada X_i for $\leq y$. Então, obtenha a fdp e $E(Y)$.]

79. Suponha que, para certo indivíduo, a ingestão de calorias no café da manhã seja uma variável aleatória com valor esperado de 500 e desvio padrão de 50, a ingestão de calorias no almoço seja aleatória com valor esperado de 900 e desvio padrão de 100, e a ingestão de calorias no jantar seja uma variável aleatória com valor esperado de 2000 e desvio padrão de 180. Assumindo que os consumos nas diferentes refeições sejam independentes, qual é a probabilidade de a ingestão média diária de calorias durante o ano seguinte (365 dias) ser de no máximo 3500? [*Sugestão*: sejam X_i, Y_i, e Z_i as três ingestões de calorias no dia i. Então, a ingestão total é dada por $\sum(X_i + Y_i + Z_i)$.]

80. O peso médio da bagagem de um passageiro da classe turística selecionado aleatoriamente, que voa entre duas cidades em uma determinada linha aérea, é 40 lb, e o desvio padrão é 10 lb. A média e o desvio padrão para um passageiro da classe executiva são 30 lb e 6 lb, respectivamente.

a. Se há 12 passageiros na classe executiva e 50 na classe turística em certo vôo, quais são o valor esperado e o desvio padrão do peso total da bagagem?

b. Se os pesos individuais da bagagem forem vas independentes e normalmente distribuídas, qual é a probabilidade de o peso total da bagagem ser de no máximo 2500 lb?

81. Vimos que, se $E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$, então, $E(X_1 + \cdots + X_n) = n\mu$. Em algumas aplicações, o número de X_is considerados não é um número fixo n, mas uma va N. Por exemplo: sejam N = o número de componentes levados a uma oficina em certo dia, e X_i o tempo de reparo do i-ésimo componente. Então, o tempo de reparo total é $X_1 + X_2 + \cdots + X_N$, a soma de um número *aleatório* de variáveis aleatórias. Quando N for independente dos X_is, pode-se mostrar que

$$E(X_1 + \cdots + X_N) = E(N) \cdot \mu$$

a. Se o número esperado de componentes levados em um dia for 10 e o tempo de conserto esperado de um componente enviado aleatoriamente for 40 min, qual é o tempo total esperado de conserto dos componentes enviados em um dia qualquer?

b. Suponha que certo tipo de componente chegue para conserto, de acordo com um processo de Poisson, com um índice de 5 por hora. O número esperado de defeitos por componente é 3,5. Qual é o valor esperado do número total de defeitos em componentes enviados para conserto durante um período de 4 horas? Certifique-se de indicar como sua resposta se desenvolve a partir do resultado geral dado.

82. Suponha que a proporção de eleitores rurais em um determinado estado que favorece certo candidato a governador seja de 0,45 e que a proporção de eleitores urbanos e suburbanos que o favorecem seja de 0,60. Se uma amostra de 200 eleitores rurais e 300 urbanos e suburbanos for obtida, qual será a probabilidade aproximada de pelo menos 250 desses eleitores favorecerem o candidato?

83. Seja μ o pH real de um composto químico. Será feita uma seqüência de n determinações independentes do pH da amostra. Suponha que o pH de cada amostra seja uma variável aleatória com valor esperado μ e desvio padrão 0,1. Quantas determinações serão necessárias se quisermos que a probabilidade da média amostral esteja entre 0,02 do pH real seja de no mínimo 0,95? Que teorema justifica seu cálculo de probabilidade?

84. Se a quantidade de refrigerante que consumo em certo dia é independente do consumo em qualquer outro dia e é normalmente distribuída com $\mu = 13$ oz e $\sigma = 2$, e se eu atualmente tenho duas caixas com seis garrafas de 16 oz, qual é a probabilidade de eu ainda ter refrigerante ao fim de duas semanas (14 dias)?

85. Consulte o Exercício 58 e suponha que os X_is sejam independentes com cada um tendo distribuição normal. Qual é a probabilidade de o volume total embarcado ser de no máximo 100.000 pé³?

86. Um aluno tem uma aula que supostamente termina às 9 h e outra que começa às 9h10. Suponha que a hora de término real da aula das 9 h seja uma va normalmente distribuída X_1 com média de 9h02 e desvio padrão de 1,5 min e que a hora de início da aula seguinte também seja uma va normalmente distribuída X_2 com média de 9h10 e desvio padrão de 1 min. Suponha também que o tempo necessário para sair de uma sala e chegar à outra seja uma va normalmente distribuída X_3 com média de 6 min e desvio padrão de 1 min. Qual é a probabilidade de o aluno chegar à segunda aula antes de ela começar? (Assuma independência de X_1, X_2, e X_3, que é razoável se o aluno não prestar atenção à hora do término da primeira aula.)

87. **a.** Use a fórmula geral da variância de uma combinação linear para escrever uma expressão para $V(aX + Y)$. Então, seja $a = \sigma_Y/\sigma_X$, e mostre que $\rho \geq -1$. [*Sugestão*: a variância sempre é ≥ 0, e $\text{Cov}(X, Y) = \sigma_X \cdot \sigma_Y \cdot \rho$.]

b. Considerando que $V(aX - Y)$, conclua que $\rho \leq 1$.

c. Use o fato de $V(W) = 0$ somente se W for uma constante para mostrar que $\rho = 1$ somente se $Y = aX + b$.

88. Suponha que a pontuação verbal X e a pontuação quantitativa Y de um indivíduo escolhido aleatoriamente em um exame de aptidão administrado nacionalmente tenham fdp conjunta.

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Você é solicitado a fornecer uma previsão t da pontuação total do indivíduo $X + Y$. O erro de previsão é o erro médio quadrado $E[(X + Y - t)^2]$. Que valor de t minimiza o erro de previsão?

89. **a.** X_1 tem distribuição qui-quadrado com parâmetro ν_1 (veja a Seção 4.4), X_2 é independente de X_1 e tem uma distribuição qui-quadrado com parâmetro ν_2. Use a técnica do Exemplo 5.21 para mostrar que $X_1 + X_2$ tem uma distribuição qui-quadrado com parâmetro $\nu_1 + \nu_2$.

b. No Exercício 65 do Capítulo 4, você foi solicitado a mostrar que, se Z é uma va normal-padrão, então Z^2 possui uma distribuição qui-quadrado com $\nu = 1$. Sejam $Z_1, Z_2, \ldots, Z_n$, as vas normais-padrão independentes de n. Qual é a distribuição de $Z_1^2 + \cdots + Z_n^2$? Justifique sua resposta.

c. Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição normal com média μ e variância σ^2. Qual é a distribuição da soma $Y = \sum_{i=1}^{n} [(X_i - \mu)/\sigma]^2$? Justifique sua resposta.

90. **a.** Mostre que $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$.

b. Sejam X_1 e X_2 as pontuações verbal e quantitativa em um exame de aptidão, e Y_1 e Y_2 as pontuações correspondentes em outro exame. Se $\text{Cov}(X_1, Y_1) = 5$, $\text{Cov}(X_1, Y_2) = 1$, $\text{Cov}(X_2, Y_1) = 2$, e $\text{Cov}(X_2, Y_2) = 8$, qual é a covariância entre as duas pontuações totais $X_1 + X_2$ e $Y_1 + Y_2$?

91. Um espécime de rocha de uma área específica é selecionado aleatoriamente e pesado duas vezes. Represente por W o peso real e por X_1 e X_2 os dois pesos medidos. Então, $X_1 = W + E_1$ e $X_2 = W + E_2$, onde E_1 e E_2 são dois erros de medição. Suponha que os E_is sejam independentes entre si e de W e que $V(E_1) = V(E_2) = \sigma_E^2$.

a. Expresse por ρ o coeficiente de correlação entre os dois pesos medidos X_1 e X_2, em termos de σ_W^2, a variância do peso real, e por σ_X^2, a variância do peso medido.

b. Calcule ρ quando $\sigma_W = 1$ kg e $\sigma_E = 0{,}01$ kg.

92. Sejam A a porcentagem de um elemento de um espécime de rocha selecionado aleatoriamente, e B a porcentagem de um segundo elemento desse mesmo espécime. Suponha que D e E sejam erros de medição ao determinar os valores de A e B, de modo que os valores medidos sejam $X = A + D$ e $Y = B + E$, respectivamente. Assuma que os erros de medição sejam independentes um do outro e dos valores reais.

a. Mostre que

$$\text{Corr}(X, Y) = \text{Corr}(A, B) \cdot \sqrt{\text{Corr}(X_1, X_2)} \cdot \sqrt{\text{Corr}(Y_1, Y_2)}$$

onde X_1 e X_2 são medições duplas do valor de A e Y_1 e Y_2 são definidos analogamente com relação a B. Que efeito a presença de erro de medição tem sobre a correlação?

b. Qual é o valor máximo de $\text{Corr}(X, Y)$ quando $\text{Corr}(X_1, X_2) = 0{,}8100$ e $\text{Corr}(Y_1, Y_2) = 0{,}9025$? Isso é preocupante?

93. Sejam $X_1, \ldots, X_n$ vas independentes com valores médios $\mu_1, \ldots, \mu_n$ e variâncias $\sigma_1^2, \ldots, \sigma_n^2$. Considere uma função $h(x_1, \ldots, x_n)$ e use-a para definir uma nova va $Y = h(X_1, \ldots, X_n)$. Sob condições bastante gerais sobre a função h, se os σ_is forem pequenos em relação aos correspondentes μ_is, pode-se mostrar que $E(Y) \approx h(\mu_1, \ldots, \mu_n)$ e

$$V(Y) \approx \left(\frac{\partial h}{\partial x_1}\right)^2 \cdot \sigma_1^2 + \cdots + \left(\frac{\partial h}{\partial x_n}\right)^2 \cdot \sigma_n^2$$

onde cada derivada parcial é avaliada em $(x_1, \ldots, x_n) = (\mu_1, \ldots, \mu_n)$. Suponha que três resistores com resistências X_1, X_2, X_3 sejam conectados em paralelo a uma bateria com voltagem X_4. Então, pela lei de Ohm, a corrente é

$$Y = X_4 \left[\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3}\right]$$

Sejam $\mu_1 = 10$ ohms, $\sigma_1 = 1{,}0$ ohm, $\mu_2 = 15$ ohms, $\sigma_2 = 1{,}0$ ohm, $\mu_3 = 20$ ohms, $\sigma_3 = 1{,}5$ ohms, $\mu_4 = 120$ V, $\sigma_4 = 4{,}0$ V. Calcule o valor esperado aproximado e o desvio padrão da corrente (sugerido por "Random Samplings" (Amostras Aleatórias) *CHEMTECH*, 1984, p. 696-697).

94. Uma aproximação mais precisa de $E[h(X_1, \ldots, X_n)]$ do Exercício 93 é

$$h(\mu_1, \ldots, \mu_n) + \frac{1}{2}\sigma_1^2\left(\frac{\partial^2 h}{\partial x_1^2}\right) + \cdots + \frac{1}{2}\sigma_n^2\left(\frac{\partial^2 h}{\partial x_n^2}\right)$$

Calcule-a para $Y = h(X_1, X_2, X_3, X_4)$, dado no Exercício 93, e compare-a ao termo principal $h(\mu_1, \ldots, \mu_n)$.

Bibliografia

LARSEN, Richard e MARX Morris, *An Introduction to Mathematical Statistics and its Applications* (3. ed.). Prentice Hall, Englewood Cliffs, NJ, 2000. Cobertura mais limitada que a do livro de Olkin et al., porém mais bem escrita e agradável.

OLKIN, Ingram, DERMAN, Cyrus e GLESER, Leon *Probability Models and Application* (2. ed.), Macmillan, Nova York, 1994. Contém uma exposição cuidadosa e abrangente das distribuições conjuntas, regras de valor esperado e teoremas de limite.

6

Estimativa Pontual

Introdução

Dado um parâmetro de interesse, como a média da população μ ou proporção da população p, o objetivo da estimativa pontual é usar uma amostra para calcular um número que represente, de certo modo, uma boa previsão do valor real do parâmetro. O número resultante é chamado estimativa pontual. Na Seção 6.1, apresentamos alguns conceitos gerais de estimativa pontual. Na Seção 6.2, descrevemos e ilustramos dois métodos importantes para se obter estimativas pontuais: o método dos momentos e o método da máxima verossimilhança.

6.1 Alguns Conceitos Gerais sobre Estimativa Pontual

A inferência estatística é quase sempre direcionada à obtenção de algum tipo de conclusão sobre um ou mais parâmetros (características da população). O processo requer que o pesquisador obtenha dados de amostras de cada população em estudo. As conclusões baseiam-se, então, nos valores calculados das várias quantidades da amostra. Por exemplo: seja μ (um parâmetro) a tensão média de quebra real das conexões do fio usado na união dos *wafers* semicondutores. Uma amostra aleatória de $n = 10$ conexões pode ser feita e a tensão de quebra de cada uma, determinada, resultando em tensões observadas $x_1, x_2, \ldots, x_{10}$. A tensão média de quebra da amostra $\bar{x}$ é, então, usada para tirar conclusões sobre o valor de μ. De forma semelhante, se σ^2 for a variância da distribuição da tensão de quebra (variância da população, outro parâmetro), o valor da variância da amostra s^2 é usado para inferir alguma coisa sobre σ^2.

Ao discutir conceitos gerais e métodos de inferência, é conveniente definir um símbolo genérico para o parâmetro de interesse. Usaremos a letra grega θ para esse fim. O objetivo da estimativa pontual é selecionar um único valor, com base nos dados da amostra, que represente um valor sensato para θ. Suponha, por exemplo, que

o parâmetro de interesse seja μ, a vida útil média real de certo tipo de pilhas. Uma amostra aleatória de $n = 3$ pilhas resulta nos valores de vida útil observados (horas) $x_1 = 5{,}0$, $x_2 = 6{,}4$, $x_3 = 5{,}9$. O valor calculado da vida útil média amostral é $\bar{x} = 5{,}77$, e é razoável considerar 5,77 um valor bastante plausível de μ — nossa "melhor previsão" para o valor de μ com base nas informações disponíveis da amostra.

Suponha que queiramos estimar um parâmetro de uma única população (por exemplo, μ ou σ) com base em uma amostra aleatória de tamanho n. Lembre-se do capítulo anterior que, antes de os dados estarem disponíveis, as observações da amostra devem ser consideradas como variáveis aleatórias (vas) $X_1, X_2, \ldots, X_n$. Acontece que qualquer função dos X_is – isto é, qualquer estatística – como a média amostral $\bar{X}$ ou o desvio padrão amostral S também é uma variável aleatória. O mesmo acontece se os dados disponíveis consistirem de mais de uma amostra. Por exemplo: podemos representar as resistências à tração de m espécimes tipo 1 e de n espécimes tipo 2 por $X_1, \ldots, X_m$ e $Y_1, \ldots, Y_n$, respectivamente. A diferença entre as duas resistências médias amostrais $\bar{X} - \bar{Y}$ é a estatística natural para fazer inferências sobre $\mu_1 - \mu_2$, a diferença entre as resistências médias da população.

DEFINIÇÃO

Uma **estimativa pontual** de um parâmetro θ é um único número que pode ser considerado um valor sensato para θ. Obtém-se uma estimativa pontual selecionando uma estatística adequada e calculando seu valor pelos dados da amostra. A estatística selecionada é chamada **estimador pontual** de θ.

No exemplo da pilha, o estimador usado para obter a estimativa pontual de μ foi $\bar{X}$, e a estimativa pontual de μ foi 5,77. Se os três valores de vida útil observados forem $x_1 = 5{,}6$, $x_2 = 4{,}5$, e $x_3 = 6{,}1$, o uso do estimador $\bar{X}$ teria resultado na estimativa $\bar{x} = (5{,}6 + 4{,}5 + 6{,}1)/3 = 5{,}40$. O símbolo $\hat{\theta}$ é habitualmente usado para representar como o estimador de θ como a estimativa pontual resultante de uma amostra dada.[1] Dessa forma, $\hat{\mu} = \bar{X}$ é lido como "estimador pontual de μ é a média amostral $\bar{X}$". A afirmativa "a estimativa pontual de μ é 5,77" é escrita resumidamente como $\hat{\mu} = 5{,}77$. Observe que, ao escrever $\hat{\theta} = 72{,}5$, não há indicação de como a estimativa pontual foi obtida (que estatística foi usada). Assim, recomenda-se que o estimador e a estimativa resultante sejam descritos.

Exemplo 6.1

Um fabricante de automóveis desenvolveu um novo tipo de pára-choque que supostamente absorve impactos com menos danos que os pára-choques anteriores. O fabricante usou pára-choques novos em uma seqüência de 25 colisões controladas contra uma parede, cada uma a 10 mph, usando um de seus modelos de carro compacto. Seja X = número de colisões que resultam em danos imperceptíveis ao automóvel. O parâmetro a ser estimado é p = proporção de todas as colisões que não provocam danos [de maneira alternativa, $p = P$(nenhum dano em uma única colisão)]. Se X observado é $x = 15$, o estimador e a estimativa mais razoáveis são

$$\text{estimador } \hat{p} = \frac{X}{n} \qquad \text{estimativa} = \frac{x}{n} = \frac{15}{25} = 0{,}60$$ ■

Se, para cada parâmetro de interesse houvesse somente um estimador pontual razoável, não haveria muito o que discutir sobre a estimativa pontual. Na maioria dos problemas, entretanto, há mais de um estimador razoável.

Exemplo 6.2

Reconsidere as 20 observações que seguem na voltagem dielétrica de quebra de peças de resina de epóxi introduzida primeiro no Exemplo 4.29 (Seção 4.6).

24,46	25,61	26,25	26,42	26,66	27,15	27,31	27,54	27,74	27,94
27,98	28,04	28,28	28,49	28,50	28,87	29,11	29,13	29,50	30,88

[1] Seguindo as notações anteriores, poderíamos usar $\hat{\Theta}$ (Θ maiúsculo) para o estimador, mas é trabalhoso escrevê-lo.

O padrão no gráfico de probabilidade normal mostrado é bastante reto; assim, assumimos que a distribuição da voltagem de quebra é normal com valor médio μ. Em virtude de as distribuições normais serem simétricas, μ também é a mediana da vida útil da distribuição. As observações dadas são, então, assumidas como o resultado de uma amostra aleatória $X_1, X_2, \ldots, X_{20}$ dessa distribuição normal. Considere os seguintes estimadores e estimativas resultantes para μ:

a. Estimador $= \overline{X}$, estimativa $= \bar{x} = \sum x_i/n = 555{,}86/20 = 27{,}793$

b. Estimador $= \tilde{X}$, estimativa $= \tilde{x} = (27{,}94 + 27{,}98)/2 = 27{,}960$

c. Estimador $= [\min(X_i) + \max(X_i)]/2 =$ a média das duas vidas úteis extremas, estimativa $= [\min(x_i) + \max(x_i)]/2 = (24{,}46 + 30{,}88)/2 = 27{,}670$

d. Estimador $= \overline{X}_{tr(10)}$, a média aparada de 10% (despreze os 10% menores e maiores da amostra e calcule a média),

$$\text{estimativa} = \bar{x}_{tr(10)} = \frac{555{,}86 - 24{,}46 - 25{,}61 - 29{,}50 - 30{,}88}{16} = 27{,}838$$

Cada um dos estimadores (a)–(d) utiliza uma medida diferente do centro da amostra para estimar μ. Qual das estimativas está mais próxima do valor real? Não podemos responder sem saber o valor real. Uma pergunta que pode ser respondida é: "Que estimador, quando usado em outras amostras dos X_is, tenderá a produzir estimativas mais próximas do valor real?" Consideraremos em breve esse tipo de pergunta. ■

Exemplo 6.3

No futuro próximo haverá um interesse crescente no desenvolvimento de ligas à base de Mg de baixo custo para diversos processos de fundição. Por isso, é importante ter maneiras práticas de determinar as várias propriedades mecânicas de tais ligas. O artigo "On the Development of a New Approach for the Determination of Yield Strength in Mg-based Alloys" (*Light Metal Age*, out. 1998, p. 50-53) propôs um método ultra-sônico para esse fim. Considere a seguinte amostra de observações no módulo de elasticidade (GPa) de espécimes da liga AZ91D de um processo de fundição:

44,2	43,9	44,7	44,2	44,0	43,8	44,6	43,1

Assuma que essas observações sejam o resultado de uma amostra aleatória $X_1, \ldots, X_8$ da distribuição da população de módulo elástico sob tais circunstâncias. Queremos estimar a variância da população σ^2. Um estimador natural é a variância amostral:

$$\hat{\sigma}^2 = S^2 = \frac{\sum(X_i - \overline{X})^2}{n-1} = \frac{\sum X_i^2 - (\sum X_i)^2/n}{n-1}$$

A estimativa correspondente é

$$\hat{\sigma}^2 = s^2 = \frac{\sum x_i^2 - (\sum x_i)^2/8}{7} = \frac{15.533{,}79 - (352{,}5)^2/8}{7} = 0{,}25125 \approx 0{,}251$$

A estimativa de σ seria, então, $\hat{\sigma} = s = \sqrt{0{,}25125} = 0{,}501$.

Um estimador alternativo seria resultado do uso do divisor n em vez de $n - 1$ (isto é, o desvio elevado ao quadrado médio):

$$\hat{\sigma}^2 = \frac{\sum(X_i - \overline{X})^2}{n} \qquad \text{estimativa} = \frac{1{,}75875}{8} = 0{,}220$$

Indicaremos em breve o porquê de muitos estatísticos preferirem S^2 ao estimador com divisor n. ■

No melhor dos mundos, podemos encontrar um estimador $\hat{\theta}$ para o qual $\hat{\theta} = \theta$, sempre. Entretanto, $\hat{\theta}$ é uma função da amostra dos X_is, portanto, é uma variável aleatória. Para algumas amostras, $\hat{\theta}$ produzirá um valor maior que θ, enquanto, para outras, $\hat{\theta}$ subestimará θ. Se escrevermos

$$\hat{\theta} = \theta + \text{erro de estimativa}$$

então, um estimador preciso seria aquele que resultasse em pequenos erros de estimativa, de modo que os valores estimados estarão próximos do valor real. Um estimador que possui as propriedades de não-tendenciosidade e variância mínima geralmente será preciso nesse sentido.

Estimadores Não-tendenciosos

Suponha que tenhamos dois instrumentos de medidas: um foi calibrado com precisão, mas o outro fornece sistematicamente leituras menores do que o valor real que está sendo medido. Quando cada instrumento for utilizado repetidas vezes no mesmo objeto devido ao erro de medição, as medições observadas não serão idênticas. Entretanto, as medições produzidas pelo primeiro instrumento serão distribuídas com relação ao valor real de tal maneira que, na média, esse instrumento mede o que se propõe a medir; assim, é chamado de instrumento não-tendencioso. O segundo instrumento produz observações que têm um componente de erro sistemático ou desvio.

DEFINIÇÃO

Um estimador pontual $\hat{\theta}$ é considerado um **estimador não-tendencioso** de θ se $E(\hat{\theta}) = \theta$ para cada valor possível *de* θ. Se $\hat{\theta}$ não for não-tendencioso, a diferença $E(\hat{\theta}) - \theta$ é chamada **desvio** ou **tendenciosidade** de $\hat{\theta}$.

Isto é, $\hat{\theta}$ é não-tendencioso se sua distribuição de probabilidade (isto é, amostragem) estiver sempre "centrada" no valor real do parâmetro. Suponha que $\hat{\theta}$ seja um estimador não-tendencioso; então, se $\theta = 100$, a distribuição amostral de $\hat{\theta}$ está centralizada em 100; se $\theta = 27{,}5$, então, a distribuição amostral de $\hat{\theta}$ está centrada em 27,5 e assim por diante. A Figura 6.1 ilustra as distribuições de diversos estimadores tendenciosos e não-tendenciosos. Observe que "centrado", aqui, significa que o valor esperado, não a mediana, da distribuição de $\hat{\theta}$ é igual a θ.

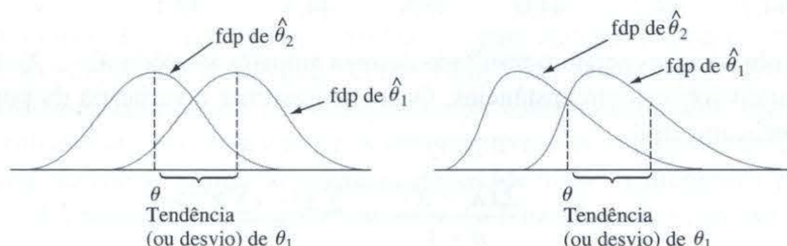


Figura 6.1 As fdps de um estimador tendencioso $\hat{\theta}_1$ e um estimador não-tendencioso θ_2, para o parâmetro θ

Talvez seja necessário saber o valor de θ (caso em que a estimativa é desnecessária) para ver se $\hat{\theta}$ é não-tendencioso. Esse geralmente não é o caso, pois muitas vezes o argumento de um valor esperado geral pode ser usado para verificar a não-tendenciosidade.

No Exemplo 6.1, a proporção amostral X/n foi usada como um estimador de p, onde X, o número de sucessos da amostra, tinha uma distribuição binomial com parâmetros n e p. Assim,

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}(np) = p$$

PROPOSIÇÃO

Quando X é uma va binomial com parâmetros n e p, a proporção amostral $\hat{p} = X/n$ é um estimador não-tendencioso de p.

Qualquer que seja o valor real de p, a distribuição do estimador $\hat{p}$ estará centrada no valor real.

Exemplo 6.4

Suponha que X, tempo de reação a certo estímulo, possua distribuição uniforme no intervalo de 0 a um limite superior desconhecido θ (assim, a função de densidade de X tem formato retangular com altura $1/\theta$ para $0 \leq x \leq \theta$). Deseja-se estimar θ com base em uma amostra aleatória $X_1, X_2, ..., X_n$ de tempos de reação. Uma vez que θ é o maior tempo possível em toda a população de tempos de reação, considere como um primeiro estimador o maior tempo de reação da amostra: $\hat{\theta}_1 = \max(X_1, ..., X_n)$. Se $n = 5$ e $x_1 = 4{,}2$, $x_2 = 1{,}7$, $x_3 = 2{,}4$, $x_4 = 3{,}9$, $x_5 = 1{,}3$, a estimativa pontual de θ é $\hat{\theta}_1 = \max(4{,}2, 1{,}7, 2{,}4, 3{,}9, 1{,}3) = 4{,}2$.

A não-tendenciosidade implica que algumas amostras produzirão estimativas que excedem θ e que outras produzirão estimativas menores que θ. Caso contrário, θ possivelmente não poderia ser o centro (ponto de equilíbrio) da distribuição dos $\hat{\theta}_1$. Entretanto, nosso estimador proposto nunca superestimará θ (o maior valor da amostra não pode exceder o maior valor da população) e subestimará θ, a menos que o maior valor amostral seja igual a θ. Este argumento intuitivo mostra que $\hat{\theta}_1$ é um estimador tendencioso. Mais precisamente, pode-se demonstrar (veja o Exercício 32) que

$$E(\hat{\theta}_1) = \frac{n}{n+1} \cdot \theta < \theta \qquad \left(\text{uma vez que } \frac{n}{n+1} < 1\right)$$

O desvio de $\hat{\theta}_1$ é dado por $n\theta/(n + 1) - \theta = -\theta/(n + 1)$, que se aproxima de 0 à medida que n aumenta. É fácil modificar $\hat{\theta}_1$ para obter um estimador não-tendencioso de θ. Considere o estimador

$$\hat{\theta}_2 = \frac{n+1}{n} \cdot \max(X_1, ..., X_n)$$

Aplicar este estimador nos dados fornece a estimativa $(6/5)(4{,}2) = 5{,}04$. O fato de $(n + 1)/n > 1$ implica que $\hat{\theta}_2$ superestimará θ para algumas amostras e o subestimará para outras. O valor médio desse estimador é

$$E(\hat{\theta}_2) = E\left[\frac{n+1}{n} \max(X_1, ..., X_n)\right] = \frac{n+1}{n} \cdot E[\max(X_1, \ldots, X_n)]$$

$$= \frac{n+1}{n} \cdot \frac{n}{n+1} \theta = \theta$$

Se $\hat{\theta}_2$ for usado repetidamente em diferentes amostras para estimar θ, algumas estimativas serão muito grandes e outras muito pequenas, mas, em geral, não haverá tendência sistemática para subestimar ou superestimar θ. ■

Princípio da Estimação Não-tendenciosa

Ao escolher dentre diversos estimadores diferentes de θ, selecione um que seja não-tendencioso.

De acordo com esse princípio, deve ser dada preferência ao estimador não-tendencioso $\hat{\theta}_2$ no Exemplo 6.4 com relação ao estimador tendencioso θ_1. Considere agora o problema de se estimar σ^2.

PROPOSIÇÃO

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória de uma distribuição com média μ e variância σ^2. Então, o estimador

$$\hat{\sigma}^2 = S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

é um estimador não-tendencioso de σ^2.

Demonstração Para qualquer va Y, $V(Y) = E(Y^2) - [E(Y)]^2$, assim, $E(Y^2) = V(Y) + [E(Y)]^2$. Aplicando essa expressão em

$$S^2 = \frac{1}{n-1}\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right]$$

fornece

$$E(S^2) = \frac{1}{n-1}\left\{\sum E(X_i^2) - \frac{1}{n}E[(\sum X_i)^2]\right\}$$

$$= \frac{1}{n-1}\left\{\sum(\sigma^2 + \mu^2) - \frac{1}{n}\{V(\sum X_i) + [E(\sum X_i)]^2\}\right\}$$

$$= \frac{1}{n-1}\left\{n\sigma^2 + n\mu^2 - \frac{1}{n}n\sigma^2 - \frac{1}{n}(n\mu)^2\right\}$$

$$= \frac{1}{n-1}\{n\sigma^2 - \sigma^2\} = \sigma^2 \quad \text{(conforme desejado)}$$ ■

O estimador que utiliza divisor n pode ser expresso como $(n-1)S^2/n$, assim,

$$E\left[\frac{(n-1)S^2}{n}\right] = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2$$

Esse estimador, portanto, é tendencioso. O desvio é $(n-1)\sigma^2/n - \sigma^2 = -\sigma^2/n$. Em virtude de o desvio ser negativo, o estimador com divisor n tende a subestimar σ^2, e é por esse motivo que o divisor $n-1$ é preferido por muitos estatísticos (mesmo quando n é grande, o desvio é pequeno e há pouca diferença entre os dois).

Embora S^2 seja não-tendencioso para σ^2, S é um estimador tendencioso de sigma (seu desvio é pequeno, a menos que n seja bem pequeno). Entretanto, existem outros bons motivos para usar S como um estimador, especialmente quando a distribuição da população é normal. Isso ficará mais visível quando discutirmos os intervalos de confiança e os testes de hipóteses nos capítulos seguintes.

No Exemplo 6.2, propusemos vários estimadores diferentes para a média μ de uma distribuição normal. Se houvesse um único estimador não-tendencioso para μ, o problema da estimação seria resolvido usando-se tal estimador. Infelizmente, esse não é o caso.

PROPOSIÇÃO

Se $X_1, X_2, \ldots, X_n$ forem uma amostra aleatória de uma distribuição com média μ, então $\overline{X}$ é um estimador não-tendencioso de μ. Se, além disso, a distribuição for contínua e simétrica, então $\tilde{X}$ e qualquer média aparada também serão estimadores não-tendenciosos de μ.

O fato de $\overline{X}$ ser não-tendencioso é simplesmente uma reafirmação de uma de nossas regras do valor esperado: $E(\overline{X}) = \mu$ para cada valor possível de μ (para distribuições discretas e contínuas). A não-tendenciosidade dos outros estimadores é mais difícil de ser verificada.

De acordo com essa proposição, apenas o princípio da não-tendenciosidade nem sempre nos permite selecionar um único estimador. Quando a população em estudo é normal, mesmo o terceiro estimador do Exemplo 6.2 é não-tendencioso, e há muitos outros estimadores não-tendenciosos. Precisamos agora de um meio de selecionar os estimadores não-tendenciosos.

Estimadores com Mínima Variância

Suponha que $\hat{\theta}_1$ e $\hat{\theta}_2$ sejam dois estimadores de θ, ambos não-tendenciosos. Então, embora a distribuição de cada estimador esteja centrada no valor real de θ, as dispersões das distribuições em torno do valor real podem ser diferentes.

Princípio da Estimação Não-tendenciosa de Mínima Variância

Dentre todos os estimadores não-tendenciosos de θ, escolha aquele que tiver variância mínima. O $\hat{\theta}$ resultante é chamado **estimador não-tendencioso de mínima variância (ENTMV)** de θ.

A Figura 6.2 ilustra as fdps de dois estimadores não-tendenciosos, com $\hat{\theta}_1$ tendo variância menor que $\hat{\theta}_2$. Então, é mais provável que $\hat{\theta}_1$ produza uma estimativa mais próxima do θ real do que $\hat{\theta}_2$. O ENTMV é, de certo modo, dentre todos os estimadores não-tendenciosos, o que produzirá, mais provavelmente, uma estimativa próxima de θ real.

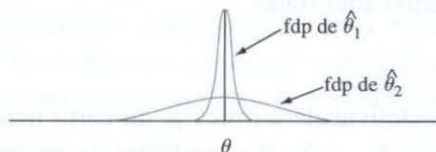


Figura 6.2 Gráficos das fdps de dois estimadores não-tendenciosos diferentes

Exemplo 6.5

Afirmamos no Exemplo 6.4 que, quando $X_1, ..., X_n$ é uma amostra aleatória de uma distribuição uniforme em $[0, \theta]$, o estimador

$$\hat{\theta}_1 = \frac{n+1}{n} \cdot \max(X_1, \ldots, X_n)$$

é não-tendencioso para θ (representamos anteriormente esse estimador por $\hat{\theta}_2$). Esse não é o único estimador não-tendencioso de θ. O valor esperado de uma va uniformemente distribuída é exatamente o ponto central do intervalo de densidade positiva, assim $E(X_i) = \theta/2$. Isso implica que $E(\overline{X}) = \theta/2$, de onde $E(2\overline{X}) = \theta$. Isto é, o estimador $\hat{\theta}_2 = 2\overline{X}$ é não-tendencioso para θ.

Se X for uniformemente distribuído no intervalo $[A, B]$, então, $V(X) = \sigma^2 = (B - A)^2/12$. Dessa forma, para o nosso caso, $V(X_i) = \theta^2/12$, $V(\overline{X}) = \sigma^2/n = \theta^2/(12n)$, e $V(\hat{\theta}_2) = V(2\overline{X}) = 4V(\overline{X}) = \theta^2/(3n)$. Os resultados do Exercício 32 podem ser usados para mostrar que $V(\hat{\theta}_1) = \theta^2/[n(n+2)]$. O estimador $\hat{\theta}_1$ tem variância menor que $\hat{\theta}_2$ se $3n < n(n+2)$—isso é, se $0 < n^2 - n = n(n-1)$. Como $n > 1$, $V(\hat{\theta}_1) < V(\hat{\theta}_2)$, então, $\hat{\theta}_1$ é um estimador melhor que θ_2. Métodos mais avançados podem ser usados para mostrar que $\hat{\theta}_1$ é o ENTMV de θ — qualquer outro estimador não-tendencioso de θ tem variância que excede $\theta^2/[n(n+2)]$. ■

Um dos triunfos da estatística matemática foi o desenvolvimento da metodologia para identificar o ENTMV em uma ampla variedade de situações. O resultado mais importante desse tipo para nossos propósitos diz respeito a estimar a média μ de uma distribuição normal.

TEOREMA Sejam $X_1, ..., X_n$ uma amostra aleatória de uma distribuição normal com parâmetros μ e σ. Então, o estimador $\hat{\mu} = \overline{X}$ é o ENTMV de μ.

Quando estamos convencidos de que a população que está sendo amostrada é normal, o resultado diz que $\overline{X}$ deve ser usado para estimar μ. No Exemplo 6.2, então, nossa estimativa seria $\overline{x} = 27{,}793$.

Em algumas situações, é possível obter um estimador com pequena tendenciosidade preferível ao melhor estimador não-tendencioso. Isso é ilustrado na Figura 6.3. Entretanto, os ENTMVs geralmente são mais fáceis de ser obtidos do que o tipo de estimador tendencioso cuja distribuição é ilustrada.

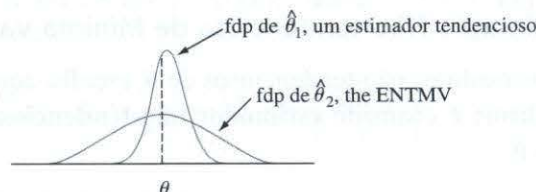


Figura 6.3 Um estimador tendencioso que é preferível ao ENTMV

Algumas Complicações

O último teorema não afirma que, ao estimar a média da população μ, o estimador $\overline{X}$ deve ser usado independentemente da distribuição que está sendo amostrada.

Exemplo 6.6

Suponha que queiramos estimar a condutividade térmica μ de certo material. Usando técnicas de medição-padrão, obteremos uma amostra aleatória $X_1, ..., X_n$ de n medições de condutividade térmica. Vamos assumir que a distribuição da população pertença a um membro de uma das três famílias a seguir:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty \tag{6.1}$$

$$f(x) = \frac{1}{\pi[1 + (x - \mu)^2]} \qquad -\infty < x < \infty \tag{6.2}$$

$$f(x) = \begin{cases} \dfrac{1}{2c} & -c \leq x - \mu \leq c \\ 0 & \text{caso contrário} \end{cases} \tag{6.3}$$

A fdp (6.1) é a distribuição normal, (6.2) é denominada distribuição de Cauchy e (6.3) é uma distribuição uniforme. As três distribuições são simétricas com relação a μ, e, de fato, a distribuição de Cauchy tem o formato de sino, mas com caudas muito mais pesadas (mais probabilidades distantes) que a curva normal. A distribuição uniforme não possui caudas. Os quatro estimadores de μ considerados anteriormente são $\overline{X}$, $\tilde{X}$, $\overline{X}_e$ (a média das duas observações extremas) e $\overline{X}_{tr(10)}$, uma média aparada.

A conclusão mais importante aqui é que o melhor estimador de μ depende crucialmente da distribuição que está sendo amostrada. Em particular,

1. Se a amostra aleatória foi tirada de uma distribuição normal, então $\overline{X}$ é o melhor dos quatro estimadores, uma vez que possui mínima variância dentre todos os estimadores não-tendenciosos.
2. Se a amostra aleatória foi tirada de uma distribuição de Cauchy, então $\overline{X}$ e $\overline{X}_e$ são péssimos estimadores de μ, enquanto $\tilde{X}$ é relativamente bom (o ENTMV não é conhecido); $\overline{X}$ é ruim porque é muito sensível às observações aberrantes, e as caudas pesadas da distribuição de Cauchy tornam provável a aparição de tais observações em qualquer amostra.
3. Se a distribuição em estudo for uniforme, o melhor estimador é $\overline{X}_e$, pois é fortemente influenciado pelas observações aberrantes; mas a ausência de caudas torna tais observações impossíveis para essa distribuição.
4. *A média aparada não é melhor em nenhuma dessas três situações, mas funciona razoavelmente bem em todas elas.* Isto é, $\overline{X}_{tr(10)}$ não sofre muito em comparação ao melhor procedimento em qualquer uma das três situações. ∎

Uma pesquisa estatística recente estabeleceu que, ao estimar um ponto de simetria μ de uma distribuição de probabilidade contínua, a média aparada com proporção de aparagem de 10% ou 20% (de cada

extremidade da amostra) produz estimativas razoavelmente comportadas sobre uma variedade bastante grande de modelos possíveis. Por essa razão, a média aparada com pequena porcentagem de aparagem é considerada um **estimador robusto.**

Em algumas situações, a escolha não está entre dois estimadores diferentes construídos a partir da mesma amostra, mas entre os estimadores baseados em dois experimentos diferentes.

Exemplo 6.7

Suponha que certo tipo de componente possua uma distribuição de vida útil exponencial com parâmetro λ, de modo que a vida útil esperada seja $\mu = 1/\lambda$. Uma amostra de n componentes é selecionada e cada um é colocado em funcionamento. Se o experimento for continuado até que todas as n vidas úteis, $X_1, \ldots, X_n$, tenham sido observadas, então $\overline{X}$ será um estimador não-tendencioso de μ.

Em alguns experimentos, entretanto, os componentes são deixados em funcionamento somente até o momento da r-ésima falha, onde $r < n$. Esse procedimento é denominado **censo**. Seja Y_1 a vida útil da primeira falha (a vida útil mínima entre os n componentes), Y_2 o tempo em que ocorre a segunda falha (a segunda menor vida útil) e assim por diante. Uma vez que o experimento termina na hora Y_r, a vida útil total acumulada no término é

$$T_r = \sum_{i=1}^{r} Y_i + (n - r)Y_r$$

Agora, demonstramos que $\hat{\mu} = T_r/r$ é um estimador não-tendencioso de μ. Para isso, precisamos de duas propriedades das variáveis exponenciais:

1. A propriedade de ausência de memória (veja a Seção 4.4), a qual diz que, em qualquer instante, a vida útil remanescente possui a mesma distribuição exponencial que a vida útil original.

2. Se $X_1, \ldots, X_k$ são independentes, cada exponencial com parâmetro λ, então, $\min(X_1, \ldots, X_k)$ é exponencial com parâmetro $k\lambda$ e possui valor esperado $1/(k\lambda)$.

Uma vez que todos os n componentes duram até Y_1, $n - 1$ dura uma quantidade de tempo adicional, $Y_2 - Y_1$, $n - 2$ uma quantidade de tempo adicional $Y_3 - Y_2$ e assim por diante, outra expressão para T_r é

$$T_r = nY_1 + (n-1)(Y_2 - Y_1) + (n-2)(Y_3 - Y_2) + \cdots + (n - r + 1)(Y_r - Y_{r-1})$$

Mas Y_1 é o mínimo de n variáveis exponenciais, assim, $E(Y_1) = 1/(n\lambda)$. De forma semelhante, $Y_2 - Y_1$ é o menor das $n - 1$ vidas úteis remanescentes, cada exponencial com parâmetro λ (pela propriedade de ausência de memória), assim, $E(Y_2 - Y_1) = 1/[(n-1)\lambda]$. Continuando, $E(Y_{i+1} - Y_i) = 1/[(n - i)\lambda]$, então,

$$\begin{aligned} E(T_r) &= nE(Y_1) + (n-1)E(Y_2 - Y_1) + \cdots + (n - r + 1)E(Y_r - Y_{r-1}) \\ &= n \cdot \frac{1}{n\lambda} + (n-1) \cdot \frac{1}{(n-1)\lambda} + \cdots + (n - r + 1) \cdot \frac{1}{(n - r + 1)\lambda} \\ &= \frac{r}{\lambda} \end{aligned}$$

Dessa forma, $E(T_r/r) = (1/r)E(T_r) = (1/r) \cdot (r/\lambda) = 1/\lambda = \mu$, conforme afirmado.

Como exemplo, suponha que 20 componentes sejam testados e $r = 10$. Então, se os primeiros 10 tempos de falha forem 11, 15, 29, 33, 35, 40, 47, 55, 58 e 72, a estimativa de μ é

$$\hat{\mu} = \frac{11 + 15 + \cdots + 72 + (10)(72)}{10} = 111{,}5$$

A vantagem do experimento com censo é que termina mais rápido que o experimento sem censo. Entretanto, pode-se mostrar que $V(T_r/r) = 1/(\lambda^2 r)$ é maior que $1/(\lambda^2 n)$, a variância de $\overline{X}$ no experimento sem censo. ■

Relatando uma Estimativa Pontual: O Erro-Padrão

Além de relatar o valor de uma estimativa pontual, deve-se indicar a sua precisão. A medida comum de precisão é o erro-padrão do estimador usado.

DEFINIÇÃO

O **erro-padrão** de um estimador θ é seu desvio padrão $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$. Se aquele envolver parâmetros desconhecidos cujos valores possam ser estimados, a substituição dessas estimativas em $\sigma_{\hat{\theta}}$ produz o **erro-padrão estimado** (desvio padrão estimado) do estimador que pode ser representado por $\hat{\sigma}_{\hat{\theta}}$ (o ^ sobre σ enfatiza que $\sigma_{\hat{\theta}}$ está sendo estimado) ou por $s_{\hat{\theta}}$.

Exemplo 6.8 (continuação do Exemplo 6.2)

Assumindo que a voltagem de quebra é normalmente distribuída, $\hat{\mu} = \overline{X}$ é o melhor estimador de μ. Se o valor de σ for conhecido como 1,5, o erro-padrão de $\overline{X}$ é $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 1{,}5/\sqrt{20} = 0{,}335$. Se, como geralmente é o caso, o valor de σ for desconhecido, a estimativa $\hat{\sigma} = s = 1{,}462$ é substituída em $\sigma_{\overline{X}}$ para se obter o erro-padrão estimado $\hat{\sigma}_{\overline{X}} = s_{\overline{X}} = s/\sqrt{n} = 1{,}462/\sqrt{20} = 0{,}327$. ■

Exemplo 6.9 (continuação do Exemplo 6.1)

O erro-padrão de $\hat{p} = X/n$ é

$$\sigma_{\hat{p}} = \sqrt{V(X/n)} = \sqrt{\frac{V(X)}{n^2}} = \sqrt{\frac{npq}{n^2}} = \sqrt{\frac{pq}{n}}$$

Uma vez que p e $q = 1 - p$ são desconhecidos, substituímos $\hat{p} = x/n$ e $\hat{q} = 1 - x/n$ em $\sigma_{\hat{p}}$, produzindo o erro-padrão estimado $\hat{\sigma}_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n} = \sqrt{(0{,}6)(0{,}4)/25} = 0{,}098$. De maneira alternativa, uma vez que o maior valor de pq é obtido quando $p = q = 0{,}5$, um limite superior no erro-padrão é $\sqrt{1/(4n)} = 0{,}10$. ■

Quando o estimador pontual $\hat{\theta}$ possui uma distribuição aproximadamente normal, que geralmente será o caso quando n for grande, então podemos ficar razoavelmente seguros de que o valor real de θ está dentro de aproximadamente 2 erros-padrão (desvios padrão) de $\hat{\theta}$. Dessa forma, se uma amostra de $n = 36$ vidas úteis de componentes fornecer $\hat{\mu} = \overline{x} = 28{,}50$ e $s = 3{,}60$, então $s/\sqrt{n} = 0{,}60$, assim, dentro de 2 erros-padrão estimados de $\hat{\mu}$, transforma-se no intervalo $28{,}50 \pm (2)(0{,}60) = (27{,}30, 29{,}70)$.

Se $\hat{\theta}$ for aproximadamente normal (não necessariamente) mas não-tendencioso, então pode-se mostrar que a estimativa desviará de θ por 4 erros-padrão no máximo 6% das vezes. Esperaríamos, então, que o valor real estivesse dentro de 4 erros-padrão de $\hat{\theta}$ (e essa é uma afirmação bastante conservadora, uma vez que se aplica a *qualquer* $\hat{\theta}$ não-tendencioso). Resumindo, o erro-padrão nos diz grosseiramente dentro de qual distância de $\hat{\theta}$ podemos esperar que o valor real de θ esteja.

A forma do estimador $\hat{\theta}$ pode ser suficientemente complicada, de modo que a teoria estatística-padrão não pode ser aplicada a fim de se obter uma expressão para $\sigma_{\hat{\theta}}$. Isso é verdade, por exemplo, no caso $\theta = \sigma$, $\hat{\theta} = S$; o desvio padrão da estatística S, σ_S, geralmente não pode ser determinado. Recentemente, um novo método computacional chamado ***bootstrap*** foi desenvolvido para tratar desse problema. Suponha que a fdp da população seja $f(x;\, \theta)$, um membro de uma família paramétrica, e que os dados $x_1, x_2, \ldots, x_n$ fornecem $\hat{\theta} = 21{,}7$. Agora, usamos o computador para obter "amostras do *bootstrap*" da fdp $f(x;\, 21.7)$ e, para cada amostra, calculamos uma "estimativa de *bootstrap*" $\hat{\theta}^*$:

Primeira amostra do *bootstrap*: $x_1^*, x_2^*, \ldots, x_n^*$; estimativa $= \hat{\theta}_1^*$

Segunda amostra do *bootstrap*: $x_1^*, x_2^*, \ldots, x_n^*$; estimativa $= \hat{\theta}_2^*$

⋮

B-ésima amostra do *bootstrap*: $x_1^*, x_2^*, \ldots, x_n^*$; estimativa $= \hat{\theta}_B^*$

Geralmente é usado $B = 100$ ou 200. Agora, seja $\overline{\theta}^* = \sum\hat{\theta}_i^*/B$ a média amostral das estimativas do *bootstrap*. A **estimativa do *bootstrap*** do erro-padrão de $\hat{\theta}$'s é agora apenas o desvio padrão amostral dos $\hat{\theta}_i^*$'s:

$$S_{\hat{\theta}} = \sqrt{\frac{1}{B-1}\sum(\hat{\theta}_i^* - \overline{\theta}^*)^2}$$

(Na literatura sobre *bootstrap*, B geralmente é usado no lugar de $B - 1$; para valores típicos de B, há, em geral, pouca diferença entre as estimativas resultantes.)

Exemplo 6.10

Um modelo teórico sugere que X, o tempo de quebra de um fluido isolante entre os elétrodos em uma voltagem específica, possui $f(x; \lambda) = \lambda e^{-\lambda x}$, uma distribuição exponencial. Uma amostra aleatória de $n = 10$ tempos de quebra (min) fornece os seguintes dados:

41,53 18,73 2,99 30,34 12,33 117,52 73,02 223,63 4,00 26,78

Uma vez que $E(X) = 1/\lambda$, $E(\overline{X}) = 1/\lambda$, então uma estimativa razoável de λ é $\hat{\lambda} = 1/\overline{x} = 1/55{,}087 = 0{,}018153$. Usamos um pacote estatístico de computador para obter $B = 100$ amostras do *bootstrap*, cada uma de tamanho 10, a partir de $f(x; 0{,}018153)$. A primeira de tais amostras foi 41,00, 109,70, 16,78, 6,31, 6,76, 5,62, 60,96, 78,81, 192,25, 27,61, da qual $\sum x_i^* = 545{,}8$ e $\hat{\lambda}_1^* = 1/54{,}58 = 0{,}01832$. A média das 100 estimativas do *bootstrap* é $\overline{\lambda}^* = 0{,}02153$, e o desvio padrão amostral dessas 100 estimativas é $s_{\hat{\lambda}} = 0{,}0091$, a estimativa do *bootstrap* do erro-padrão dos $\hat{\lambda}$'s. Um histograma dos 100 $\hat{\lambda}_i^*$'s estava positivamente inclinado, sugerindo que a distribuição amostral de $\hat{\lambda}$ também possuía essa propriedade. ■

Às vezes, o pesquisador deseja estimar a característica de uma população sem assumir que sua distribuição pertence a uma família paramétrica particular. Um caso desses ocorreu no Exemplo 6.6, em que uma média aparada 10% foi proposta para estimar o centro da distribuição simétrica da população θ. Os dados do Exemplo 6.2 forneceram $\hat{\theta} = \overline{x}_{tr(10)} = 27{,}838$, mas agora não há $f(x; \theta)$ assumido. Então, como podemos obter uma amostra do *bootstrap*? A resposta é considerar a própria amostra como constituinte da população (as $n = 10$ observações no Exemplo 6.2) e tirar B amostras diferentes, cada uma de tamanho n, com substituição da referida população. O livro de Bradley Efron e Robert Tibshirani ou o de John Rice, listados na bibliografia, fornecem mais informações.

Exercícios | Seção 6.1 (1–19)

1. Os dados que seguem sobre a resistência à flexão (MPa) das vigas de um tipo de concreto foram introduzidos no Exemplo 1.2.

5,9	7,2	7,3	6,3	8,1	6,8	7,0
7,6	6,8	6,5	7,0	6,3	7,9	9,0
8,2	8,7	7,8	9,7	7,4	7,7	9,7
7,8	7,7	11,6	11,3	11,8	10,7	

a. Calcule uma estimativa pontual do valor médio da resistência para a população conceitual de todas as vigas fabricadas dessa forma e diga qual estimador você utilizou. (*Sugestão*: $\sum x_i = 219{,}8$.)

b. Calcule uma estimativa pontual do valor da resistência que separa as 50% mais fracas de todas as vigas das 50% mais fortes e diga qual estimador você utilizou.

c. Calcule e interprete uma estimativa pontual do desvio padrão da população σ. Que estimador você utilizou? (*Sugestão*: $\sum x_i^2 = 1860{,}94$.)

d. Calcule uma estimativa pontual da proporção de todas as vigas cuja resistência à flexão exceda 10 MPa. (*Sugestão*: considere como observação um "sucesso" se exceder 10.)

e. Calcule uma estimativa pontual do coeficiente de variação σ/μ da população e diga qual estimador você utilizou.

2. Uma amostra de 20 alunos que recentemente tiveram estatística básica descobriu as seguintes informações sobre a marca das suas calculadoras (T = Texas Instruments, H = Hewlett Packard, C = Cassio, S = Sharp)

T T H T C T T S C H

S S T H C T T T H T

a. Calcule a proporção real de todos os alunos que possuem uma calculadora Texas Instruments.

b. Dos 10 alunos que tinham uma calculadora TI, 4 possuíam calculadoras com gráfico. Calcule a proporção de alunos que não possuem uma calculadora com gráfico TI.

3. Considere a seguinte amostra de observações sobre a espessura da camada da pintura de baixa viscosidade ("Achieving a Target Value for a Manufacturing Process: A Case Study," *J. of Quality Technology,* 1992, p. 22-26):

0,83	0,88	0,88	1,04	1,09	1,12	1,29	1,31
1,48	1,49	1,59	1,62	1,65	1,71	1,76	1,83

Assuma que a distribuição da espessura da camada seja normal (um gráfico de probabilidade normal suporta fortemente essa hipótese).

a. Calcule uma estimativa pontual do valor médio da espessura da camada e diga qual estimador você utilizou.

b. Calcule uma estimativa pontual da mediana da distribuição da espessura da camada e diga qual estimador você utilizou.

c. Calcule uma estimativa pontual do valor que separa os 10% maiores de todos os valores na distribuição da espessura dos 90% remanescentes e diga qual estimador você utilizou. (*Sugestão*: mostre o que você está tentando estimar em termos de μ e σ.)

d. Calcule $P(X < 1,5)$, isto é, a proporção de todos os valores da espessura menores que 1,5. (*Sugestão*: Se você conhecesse os valores de μ e σ, poderia calcular essa probabilidade. Esses valores não estão disponíveis, mas podem ser estimados.)

e. Qual é o erro-padrão estimado do estimador que você utilizou na parte (b)?

4. O artigo do qual os dados do Exercício 1 foram tirados também forneceu as observações de resistência a seguir para os cilindros:

6,1	5,8	7,8	7,1	7,2	9,2	6,6	8,3	7,0	8,3
7,8	8,1	7,4	8,5	8,9	9,8	9,7	14,1	12,6	11,2

Antes de obter os dados, represente as resistências da viga por $X_1, \ldots, X_m$ e as resistências do cilindro por $Y_1, \ldots, Y_n$. Suponha que os X_is constituem uma amostra aleatória de uma distribuição com média μ_1 e desvio padrão σ_1 e que os Y_is formem uma amostra aleatória (independente dos X_is) de outra distribuição com média μ_2 e desvio padrão σ_2.

a. Use as regras do valor esperado para mostrar que $\overline{X} - \overline{Y}$ é um estimador não-tendencioso de $\mu_1 - \mu_2$. Calcule a estimativa para os dados fornecidos.

b. Use as regras de variância do Capítulo 5 para obter uma expressão para a variância e o desvio padrão (erro-padrão) do estimador da parte (a) e calcule o erro-padrão estimado.

c. Calcule uma estimativa pontual da relação σ_1/σ_2 dos dois desvios padrão.

d. Suponha que uma única viga e um único cilindro sejam selecionados aleatoriamente. Calcule uma estimativa pontual da variância da diferença $X - Y$ entre a resistência da viga e a resistência do cilindro.

5. Como exemplo de uma situação em que várias estatísticas diferentes podiam razoavelmente ser usadas para calcular uma estimativa pontual, considere uma população de N faturas. Associado a cada fatura está seu "valor contábil", o valor registrado dessa fatura. Seja por T o valor contábil total, um valor conhecido. Alguns desses valores são errôneos. Uma auditoria será realizada, selecionando aleatoriamente n faturas e determinando o valor (correto) examinado de cada uma. Suponha que a amostra forneça os seguintes resultados (em dólares).

	Fatura				
	1	**2**	**3**	**4**	**5**
Valor contábil	300	720	526	200	127
Valor examinado	300	520	526	200	157
Erro	0	200	0	0	−30

Sejam

$\overline{Y}$ = valor contábil médio amostral

$\overline{X}$ = valor médio amostral examinado

$\overline{D}$ = erro médio amostral

Diversas estatísticas diferentes para estimar o valor (correto) examinado total foram propostas (veja "Statistical Models and Analysis in Auditing", *Statistical Science,* 1989, p. 2-33). Elas incluem

$$\text{Média por estatística da unidade} = N\overline{X}$$

$$\text{Estatística da diferença} = T - N\overline{D}$$

$$\text{Estatística da razão} = T \cdot (\overline{X}/\overline{Y})$$

Se $N = 5000$ e $T = 1.761,300$, calcule as três estimativas pontuais correspondentes. (O artigo mencionado discute as propriedades desses estimadores.)

6. Considere as observações que seguem sobre o fluxo direto (1000 acre-pés) registrado em um posto em Colorado de 1º de abril a 31 de agosto, por um período de 31 anos (Dados extraídos de um artigo do volume de 1974 de *Water Resources Research*).

127,96	210,07	203,24	108,91	178,21
285,37	100,85	89,59	185,36	126,94
200,19	66,24	247,11	299,87	109,64
125,86	114,79	109,11	330,33	85,54
117,64	302,74	280,55	145,11	95,36
204,91	311,13	150,58	262,09	477,08
94,33				

Um gráfico de probabilidade apropriado confirma o uso da distribuição lognormal (veja Seção 4.5) como um modelo razoável do fluxo dos rios.

a. Calcule os parâmetros da distribuição. [*Sugestão*: lembre-se que X possui distribuição lognormal com parâmetros μ e σ^2 se $\ln(X)$ for normalmente distribuído com média μ e variância σ^2.]

b. Use as estimativas da parte (a) para calcular uma estimativa do valor esperado de fluxo dos rios. [*Sugestão*: qual é o $E(X)$?]

7. a. Uma amostra aleatória de 10 casas em uma área específica, todas aquecidas com gás natural, é selecionada; a quantidade de gás (termos) usada durante o mês de janeiro é determinada para cada casa. As observações resultantes são: 103, 156, 118, 89, 125, 147, 122, 109, 138, 99. Seja μ o consumo médio de gás durante janeiro por todas as casas dessa área. Calcule a estimativa pontual de μ.

b. Suponha que existam 10.000 casas nessa área que utilizem gás natural para aquecimento. Seja τ a quantidade total de gás usado por todas essas casas no mês de janeiro. Calcule τ, usando os dados da parte (a). Que estimador você utilizou ao calcular sua estimativa?

c. Use os dados na parte (a) para calcular p, a proporção de todas as casas que usaram no mínimo 100 termos.

d. Forneça uma estimativa pontual do uso mediano da população (o valor médio na população de todas as casas) com base na amostra da parte (a). Que estimador você utilizou?

8. Em uma amostra aleatória de certo tipo de 80 componentes, 12 são considerados defeituosos.

a. Forneça uma estimativa pontual da proporção de todos os componentes que *não* são defeituosos.

b. Um sistema será construído, selecionando-se aleatoriamente dois desses componentes e conectando-os em série, conforme ilustramos abaixo:



A conexão em série implica que o sistema funcionará somente se nenhum componente estiver com defeito (isto é, se ambos funcionarem adequadamente). Calcule a proporção de todos os sistemas que funcionam adequadamente. [*Sugestão*: se p representa a probabilidade de um componente funcionar corretamente, como pode P (funcionamentos do sistema) ser expresso em termos de p?]

9. Todos os 150 itens fabricados recentemente são examinados e o número de arranhões por item é registrado (os itens são supostamente livres de arranhões), produzindo os seguintes dados:

Número de arranhões por item	0	1	2	3	4	5	6	7
Freqüência observada	18	37	42	30	13	7	2	1

Seja X = o número de arranhões em um item escolhido aleatoriamente, assuma que X possui uma distribuição de Poisson com parâmetro λ.

a. Determine um estimador não-tendencioso de λ e calcule a estimativa dos dados. [*Sugestão*: $E(X) = \lambda$ para Poisson de X, então $E(\overline{X}) = ?$]

b. Qual é o desvio padrão (erro-padrão) do seu estimador? Calcule o erro-padrão estimado. (*Sugestão*: $\sigma_X^2 = \lambda$ para Poisson de X.)

10. Usando um bastão longo de comprimento μ, você traçará um gráfico quadrado em que o comprimento de cada lado é μ. Assim, a área do gráfico será μ^2. Entretanto, você não conhece o valor de μ, então decide fazer n medições independentes $X_1, X_2, ..., X_n$ do comprimento. Assuma que cada X_i tenha média μ (medições não-tendenciosas) e variância σ^2.

a. Mostre que $\overline{X}^2$ não é um estimador não-tendencioso para μ^2. [*Sugestão*: para qualquer va Y, $E(Y^2) = V(Y) + [E(Y)]^2$. Aplique a expressão com $Y = \overline{X}$.]

b. Para que valor de k o estimador $\overline{X}^2 - kS^2$ é não-tendencioso para μ^2? [*Sugestão*: calcule $E(\overline{X}^2 - kS^2)$.]

11. De n_1 fumantes do sexo masculino escolhidos aleatoriamente, X_1 fumavam cigarro com filtro, enquanto de n_2 fumantes do sexo feminino selecionadas aleatoriamente, X_2 fumavam cigarro com filtro. Represente por p_1 e p_2 as probabilidades de um homem e uma mulher selecionados de modo aleatório, respectivamente, fumarem cigarro com filtro.

a. Mostre que $(X_1/n_1) - (X_2/n_2)$ é um estimador não-tendencioso para $p_1 - p_2$. [*Sugestão*: $E(X_i) = n_i p_i$ para $i = 1, 2$.]

b. Qual é o erro-padrão do estimador da parte (a)?

c. Como você usaria os valores observados de x_1 e x_2 para calcular o erro-padrão do seu estimador?

d. Se $n_1 = n_2 = 200$, $x_1 = 127$ e $x_2 = 176$, use o estimador da parte (a) para obter uma estimativa de $p_1 - p_2$.

e. Use o resultado da parte (c) e os dados da parte (d) para calcular o erro-padrão do estimador.

12. Suponha que certo tipo de fertilizante possua um rendimento esperado por acre de μ_1, com variância σ^2, enquanto o rendimento esperado de um segundo tipo de fertilizante seja μ_2 com a mesma variância σ^2. Represente por S_1^2 e S_2^2 as variâncias amostrais dos rendimentos com base nos tamanhos de amostra n_1 e n_2, respectivamente, dos dois fertilizantes. Mostre que o estimador combinado

$$\hat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

é um estimador não-tendencioso de σ^2.

13. Considere uma amostra aleatória $X_1, ..., X_n$ da fdp

$$f(x; \theta) = 0{,}5(1 + \theta x) \qquad -1 \leq x \leq 1$$

em que $-1 \le \theta \le 1$ (essa distribuição surge na física de partículas). Mostre que $\hat{\theta} = 3\overline{X}$ é um estimador não-tendencioso de θ. [*Sugestão*: determine primeiro $\mu = E(X) = E(\overline{X})$.]

14. Uma amostra de n aeronaves *Pandemonium jet fighters* capturadas resulta em números de série $x_1, x_2, x_3, \ldots, x_n$. A CIA sabe que cada aeronave foi numerada consecutivamente na fábrica, começando com α e terminando com β, de modo que o número total de aviões fabricados é $\beta - \alpha + 1$ (por exemplo: se $\alpha = 17$ e $\beta = 29$, então, $29 - 17 + 1 = 13$ aviões que têm números de série 17, 18, 19, ..., 28, 29, foram fabricados). Entretanto, a CIA não conhece os valores de α ou β. Um estatístico dessa instituição sugere o uso do estimador $\max(X_i) - \min(X_i) + 1$ para calcular o número total de aviões fabricados.

a. Se $n = 5$, $x_1 = 237$, $x_2 = 375$, $x_3 = 202$, $x_4 = 525$ e $x_5 = 418$, qual é a estimativa correspondente?

b. Sob quais condições na amostra o valor da estimativa será exatamente igual ao número total real de aviões? A estimativa sempre será maior que o total real? Você acha que o estimador é não-tendencioso para calcular $\beta - \alpha + 1$? Explique em uma ou duas frases.

15. Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma distribuição de Rayleigh com fdp

$$f(x;\,\theta) = \frac{x}{\theta}\,e^{-x^2/(2\theta)} \qquad x > 0$$

a. Pode-se mostrar que $E(X^2) = 2\theta$. Use esse fato para construir um estimador não-tendencioso de θ com base em $\sum X_i^2$ (use regras de valor esperado para mostrar que é não-tendencioso).

b. Calcule θ a partir das $n = 10$ observações a seguir sobre a força vibratória de uma palheta de turbina sob condições especificadas:

16,88	10,23	4,59	6,66	13,68
14,23	19,87	9,40	6,51	10,95

16. Suponha que o crescimento médio real μ de um tipo de planta durante o período de um ano seja idêntico ao de um segundo tipo, mas a variância de crescimento do primeiro tipo seja σ^2, enquanto a do segundo é $4\sigma^2$. Sejam $X_1, \ldots, X_m$, m observações independentes de crescimento para o primeiro tipo [assim, $E(X_i) = \mu$, $V(X_i) = \sigma^2$], e $Y_1, \ldots, Y_n$, n observações independentes de crescimento para o segundo tipo [$E(Y_i) = \mu$, $V(Y_i) = 4\sigma^2$].

a. Mostre que, para qualquer δ entre 0 e 1, o estimador $\hat{\mu} = \delta\overline{X} + (1 - \delta)\overline{Y}$ é não-tendencioso para μ.

b. Para m e n fixos, calcule $V(\hat{\mu})$ e depois determine o valor de δ que minimiza $V(\hat{\mu})$. [*Sugestão*: diferencie $V(\hat{\mu})$ em relação a δ.]

17. No Capítulo 3, definimos uma va binomial negativa como o número de falhas que ocorrem antes do r-ésimo sucesso em uma seqüência de tentativas de sucesso / falha idênticas e independentes. A função de massa de probabilidade (fmp) de X é

$$nb(x;\,r,\,p) = \begin{cases} \binom{x + r - 1}{x} p^r (1-p)^x & x = 0, 1, 2, \ldots \\ 0 & \text{caso contrário} \end{cases}$$

a. Suponha que $r \ge 2$. Mostre que

$$\hat{p} = (r - 1)/(X + r - 1)$$

é um estimador não-tendencioso para p. [*Sugestão*: Desenvolva $E(\hat{p})$ e cancele $x + r - 1$ dentro da soma.]

b. Um repórter que deseja entrevistar cinco pessoas que apóiam um candidato começa perguntando a elas se apóiam (S) ou não (F) tal candidato. Se a seqüência das respostas é $SFFSFFFSSS$, estime p = proporção real que apóia o candidato.

18. Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma fdp $f(x)$ simétrica em relação a μ, de modo que $\tilde{X}$ seja um estimador não-tendencioso de μ. Se n for grande, pode-se mostrar que $V(\tilde{X}) \approx 1/(4n[f(\mu)]^2)$.

a. Compare $V(\tilde{X})$ a $V(\overline{X})$ quando a distribuição subjacente for normal.

b. Quando a fdp subjacente for Cauchy (veja o Exemplo 6.6), $V(\overline{X}) = \infty$, então $\overline{X}$ será um terrível estimador. Qual é $V(\tilde{X})$, neste caso, quando n for grande?

19. Uma pesquisadora deseja estimar a proporção de alunos que violaram o código de honra em certa universidade. Ao obter uma amostra aleatória de n alunos, ela percebe que perguntar a cada um: "Você violou o código de honra?" provavelmente resultará em algumas respostas falsas. Considere o seguinte esquema, chamado de técnica de **resposta aleatória**. A pesquisadora monta um baralho de 100 cartas, das quais 50 são do tipo I e 50 do tipo II.

Tipo I: Você violou o código de honra (sim ou não)?

Tipo II: O último dígito de seu número de telefone é 0, 1 ou 2 (sim ou não)?

Cada aluno da amostra aleatória é solicitado a misturar o baralho, tirar uma carta e responder à pergunta resultante com sinceridade. Devido à pergunta irrelevante nas cartas do tipo II, a resposta sim não estigmatiza mais o entrevistado, de modo que assumimos que as respostas sejam verdadeiras. Seja p a proporção dos violadores do código de honra (isto é, a probabilidade de um aluno selecionado aleatoriamente ser um violador) e por $\lambda = P$ (resposta sim). Então, λ e p são relacionados por $1\lambda = 0{,}5p + (0{,}5)(0{,}3)$.

a. Seja Y o número de respostas afirmativas, de modo que $Y \sim \text{Bin}(n, \lambda)$. Dessa forma, Y/n é um estimador não-tendencioso de λ. Deduza um estimador para p com base em Y. Se $n = 80$ e $y = 20$, qual é sua estimativa? (*Sugestão*: resolva $\lambda = 0{,}5p + 0{,}15$ para p e então substitua Y/n por λ.)

b. Use o fato de $E(Y/n) = \lambda$ para mostrar que seu estimador $\hat{p}$ é não-tendencioso.

c. Se havia 70 cartas tipo I e 30 tipo II, qual seria seu estimador para p?

6.2 Métodos de Estimativa Pontual

A definição de não-tendenciosidade geralmente não indica como os estimadores não-tendenciosos podem ser deduzidos. Agora, discutimos dois métodos "construtivos" para obter estimadores pontuais: o método dos momentos e o método da máxima verossimilhança. Por "construtivo" queremos dizer que a definição geral de cada tipo de estimador sugere explicitamente como obter o estimador em qualquer problema específico. Embora os estimadores da máxima verossimilhança sejam geralmente preferíveis aos estimadores de momentos devido a certas propriedades de eficiência, eles usualmente precisam significativamente de mais cálculos que os estimadores de momentos. Às vezes, é o caso de esses métodos produzirem estimadores não-tendenciosos.

O Método dos Momentos

A idéia básica deste método é igualar certas características da amostra, como a média, aos valores esperados correspondentes da população. Então, resolvendo essas equações para os valores dos parâmetros desconhecidos produz os estimadores.

DEFINIÇÃO

Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma fmp ou fdp $f(x)$. Para $k = 1, 2, 3, \ldots$, o **k-ésimo momento da população** ou **k-ésimo momento da distribuição $f(x)$** é $E(X^k)$. O **k-ésimo momento amostral** é $(1/n)\sum_{i=1}^{n} X_i^k$.

Dessa forma, o primeiro momento da população é $E(X) = \mu$ e o primeiro momento amostral é $\sum X_i/n = \overline{X}$. Os segundos momentos da população e da amostra são $E(X^2)$ e $\sum X_i^2/n$, respectivamente. Os momentos da população serão funções de quaisquer parâmetros desconhecidos $\theta_1, \theta_2, \ldots$

DEFINIÇÃO

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma distribuição com fmp ou fdp $f(x; \theta_1, \ldots, \theta_m)$, onde $\theta_1, \ldots, \theta_m$ são parâmetros cujos valores são desconhecidos. Então, os **estimadores pelo método dos momentos** $\hat{\theta}_1, \ldots, \hat{\theta}_m$ são obtidos igualando-se os primeiros m momentos da amostra aos primeiros m momentos da população correspondente e resolvendo para $\theta_1, \ldots, \theta_m$.

Se, por exemplo, $m = 2$, $E(X)$ e $E(X^2)$ serão funções de θ_1 e θ_2. Definindo $E(X) = (1/n)\sum X_i$ $(= \overline{X})$ e $E(X^2) = (1/n)\sum X_i^2$ fornece duas equações em θ_1 e θ_2. A solução, então, define os estimadores. Para estimar a média da população μ, o método fornece $\mu = \overline{X}$, assim, o estimador é a média amostral.

Exemplo 6.11

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de tempos de serviço de n clientes em uma determinada instalação, onde a distribuição subjacente é considerada exponencial com parâmetro λ. Uma vez que há somente um parâmetro a ser estimado, o estimador é obtido igualando $E(X)$ a $\overline{X}$. Desde que $E(X) = 1/\lambda$ para uma distribuição exponencial, temos $1/\lambda = \overline{X}$ ou $\lambda = 1/\overline{X}$. O estimador pelo método dos momentos de λ é, então, $\lambda = 1/\overline{X}$. ■

Exemplo 6.12

Seja $X_1, \ldots, X_n$, uma amostra aleatória de uma distribuição gama com parâmetros α e β. Pela Seção 4.4, $E(X) = \alpha\beta$ e $E(X^2) = \beta^2\Gamma(\alpha + 2)/\Gamma(\alpha) = \beta^2(\alpha + 1)\alpha$. Os estimadores pelo método dos momentos de α e β são obtidos, resolvendo-se

$$\overline{X} = \alpha\beta \qquad \frac{1}{n}\sum X_i^2 = \alpha(\alpha + 1)\beta^2$$

Uma vez que $\alpha(\alpha + 1)\beta^2 = \alpha^2\beta^2 + \alpha\beta^2$ e a primeira equação implica $\alpha^2\beta^2 = \overline{X}^2$, a segunda equação se torna

$$\frac{1}{n}\sum X_i^2 = \overline{X}^2 + \alpha\beta^2$$

Dividindo-se cada membro dessa segunda equação pelo lado correspondente da primeira e substituindo-se de volta, são fornecidos os estimadores

$$\hat{\alpha} = \frac{\overline{X}^2}{(1/n)\sum X_i^2 - \overline{X}^2} \qquad \hat{\beta} = \frac{(1/n)\sum X_i^2 - \overline{X}^2}{\overline{X}}$$

Para ilustrar, os dados do tempo de sobrevida mencionados no Exemplo 4.21 são

152	115	109	94	88	137	152	77	160	165
125	40	128	123	136	101	62	153	83	69

com $\bar{x} = 113{,}5$ e $(1/20)\sum x_i^2 = 14{,}087.8$. As estimativas são

$$\hat{\alpha} = \frac{(113{,}5)^2}{14.087{,}8 - (113{,}5)^2} = 10.7 \qquad \hat{\beta} = \frac{14.087{,}8 - (113{,}5)^2}{113{,}5} = 10{,}6$$

Essas estimativas de α e β diferem dos valores sugeridos por Gross e Clark porque eles usaram uma técnica de estimativa diferente. ■

Exemplo 6.13

Seja $X_1, ..., X_n$ uma amostra aleatória de uma distribuição binomial negativa geral com parâmetros r e p (Seção 3.5). Uma vez que $E(X) = r(1 - p)/p$ e $V(X) = r(1 - p)/p^2$, $E(X^2) = V(X) + [E(X)]^2 = r(1 - p)(r - rp + 1)/p^2$. Igualando $E(X)$ a $\overline{X}$ e $E(X^2)$ a $(1/n)\sum X_i^2$ finalmente fornece

$$\hat{p} = \frac{\overline{X}}{(1/n)\sum X_i^2 - \overline{X}^2} \qquad \hat{r} = \frac{\overline{X}^2}{(1/n)\sum X_i^2 - \overline{X}^2 - \overline{X}}$$

Como ilustração, Reep, Pollard e Benjamin ("Skill and Chance in Ball Games," *J. Royal Stat. Soc.*, 1971, p. 623-629) consideram a distribuição binomial negativa como um modelo para o número de gols por jogo marcados pelos times da Liga Nacional de Hóquei. Os dados de 1966-1967 aparecem a seguir (420 jogos):

Gols	0	1	2	3	4	5	6	7	8	9	10
Freqüência	29	71	82	89	65	45	24	7	4	1	3

Então,

$$\bar{x} = \sum x_i/420 = [(0)(29) + (1)(71) + \cdots + (10)(3)]/420 = 2{,}98$$

e

$$\sum x_i^2/420 = [(0)^2(29) + (1)^2(71) + \cdots + (10)^2(3)]/420 = 12{,}40$$

Dessa forma,

$$\hat{p} = \frac{2{,}98}{12{,}40 - (2{,}98)^2} = 0{,}85 \qquad \hat{r} = \frac{(2{,}98)^2}{12{,}40 - (2{,}98)^2 - 2{,}98} = 16{,}5$$

Embora r por definição deva ser positivo, o denominador de $\hat{r}$ pode ser negativo, indicando que a distribuição binomial negativa não é apropriada (ou que o estimador pelo método dos momentos é inadequado). ■

Estimadores de Máxima Verossimilhança

O método da máxima verossimilhança foi introduzido primeiro por R. A. Fisher, um geneticista e estatístico, na década de 1920. A maioria dos estatísticos recomenda esse método, pelo menos quando o tamanho da amostra for grande, uma vez que os estimadores resultantes têm certas propriedades de eficiência desejáveis (veja a segunda proposição da página 240).

Exemplo 6.14

Uma amostra de 10 capacetes de ciclista fabricados por certa empresa é obtida. Em teste, descobre-se que o primeiro, o terceiro e o décimo capacetes estão com defeito, enquanto os demais não. Seja $p = P$(capacete com defeito) e defina $X_1, ..., X_{10}$ por $X_i = 1$ se o i-ésimo capacete estiver com defeito e por zero, caso contrário. Então, os x_is observados são 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, de modo que a fmp conjunta da amostra é

$$f(x_1, x_2, \ldots, x_{10}; p) = p(1-p)p \cdots p = p^3(1-p)^7 \qquad (6.4)$$

Perguntamos agora: "Para que valor de p a amostra observada é mais provável de ter ocorrido?" Isto é, queremos determinar o valor de p que maximiza a fmp (6.4) ou, de forma equivalente, que maximiza o log natural de (6.4).[2] Uma vez que

$$\ln[f(x_1, \ldots, x_{10}; p)] = 3\ln(p) + 7\ln(1-p) \qquad (6.5)$$

a qual é uma função diferenciável de p, igualando a derivada de (6.5) a zero, resultando o valor máximo[3]

$$\frac{d}{dp}\ln[f(x_1,..., x_{10}; p)] = \frac{3}{p} - \frac{7}{1-p} = 0 \quad p = \frac{3}{10} = \frac{x}{n}$$

onde x é o número observado de sucessos (capacetes com defeito). A estimativa de p é agora $\hat{p} = \frac{3}{10}$. Chama-se estimativa da máxima verossimilhança, pois, para valores fixos de $x_1, ..., x_{10}$, esse é o valor do parâmetro que maximiza a verossimilhança (fmp conjunta) da amostra observada.

Observe que, se nos dissessem somente que entre os 10 capacetes havia três que estavam com defeito, a Equação (6.4) seria substituída pela fmp binomial $\binom{10}{3}p^3(1-p)^7$, que também é maximizada para $\hat{p} = \frac{3}{10}$. ■

DEFINIÇÃO Sejam $X_1, X_2, ..., X_n$ a fmp ou fdp conjunta

$$f(x_1, x_2, ..., x_n; \theta_1, ..., \theta_m) \qquad (6.6)$$

onde os parâmetros $\theta_1, ..., \theta_m$ têm valores desconhecidos. Quando $x_1, ..., x_n$ são os valores observados da amostra e (6.6) é considerada uma função de $\theta_1, ..., \theta_m$, é denomina-se **função de verossimilhança.** As estimativas de máxima verossimilhança (emvs) $\hat{\theta}_1, ..., \hat{\theta}_m$ são os valores dos θ_is que maximizam a função de verossimilhança, de modo que

$$f(x_1, ..., x_n; \hat{\theta}_1, ..., \hat{\theta}_m) \geq f(x_1, ..., x_n; \theta_1, ..., \theta_m) \quad \text{para todos os } \theta_1, ..., \theta_m$$

Quando os X_is são substituídos no lugar de x_is, resultam os **estimadores de máxima verossimilhança**.

[2] Uma vez que $\ln[g(x)]$ é uma função monotônica de $g(x)$, determinar x que maximiza $\ln[g(x)]$ é equivalente a maximizar a própria $g(x)$. Em estatística, calcular o logaritmo geralmente muda um produto para uma soma, com a qual é mais fácil trabalhar.

[3] Esta conclusão requer a verificação da segunda derivada, mas os detalhes são omitidos.

A função verossimilhança nos diz quão provável a amostra observada é uma função dos valores dos parâmetros possíveis. Maximizar a verossimilhança fornece os valores dos parâmetros para os quais a amostra observada é mais provável de ter sido gerada – isto é, os valores de parâmetro que "mais aproximadamente concordam" com os dados observados.

Exemplo 6.15

Suponha que $X_1, X_2, \ldots, X_n$ seja a amostra aleatória de uma distribuição exponencial com parâmetro λ. Devido à independência, a função verossimilhança é o produto das fdps individuais:

$$f(x_1, \ldots, x_n; \lambda) = (\lambda e^{-\lambda x_1}) \cdot \cdots \cdot (\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda \Sigma x_i}$$

O ln(verossimilhança) é

$$\ln[f(x_1, \ldots, x_n; \lambda)] = n \ln(\lambda) - \lambda \sum x_i$$

Igualar $(d/d\lambda)$[ln(verossimilhança)] a zero resulta $n/\lambda - \sum x_i = 0$, ou $\lambda = n/\sum x_i = 1/\bar{x}$. Dessa forma, a emv é $\hat{\lambda} = 1/\bar{X}$; é idêntica ao método de estimadores pelos momentos [mas não é um estimador não-tendencioso, uma vez que $E(1/\bar{X}) \neq 1/E(\bar{X})$]. ∎

Exemplo 6.16

Seja $X_1, \ldots, X_n$ a amostra aleatória de uma distribuição normal. A função verossimilhança é

$$f(x_1, \ldots, x_n; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\mu)^2/(2\sigma^2)} \cdot \cdots \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\mu)^2/(2\sigma^2)}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\Sigma(x_i-\mu)^2/(2\sigma^2)}$$

então

$$\ln[f(x_1, \ldots, x_n; \mu, \sigma^2)] = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i - \mu)^2$$

Para determinar os valores máximos de μ e σ^2, devemos calcular as derivadas parciais de ln(f) em relação a μ e σ^2, igualá-las a zero e resolver as duas equações resultantes. Omitindo os detalhes, as emvs resultantes são

$$\hat{\mu} = \bar{X} \qquad \hat{\sigma}^2 = \frac{\sum(X_i - \bar{X})^2}{n}$$

A emv de σ^2 não é o estimador não-tendencioso; então, dois princípios diferentes de estimativa (não-tendenciosidade e máxima verossimilhança) produzem dois estimadores diferentes. ∎

Exemplo 6.17

No Capítulo 3, discutimos o uso da distribuição de Poisson para modelar o número de "eventos" que ocorrem em uma região bidimensional. Assuma que, quando a região R, que está servindo como amostra, possui área $a(R)$, o número X de eventos que ocorrem em R tem distribuição de Poisson com parâmetro $\lambda a(R)$ (em que λ é o número esperado de eventos por unidade de área) que as regiões não-sobrepostas produzem Xs independentes.

Suponha que um ecologista selecione n regiões não-sobrepostas $R_1, \ldots, R_n$ e conte o número de certa espécie de plantas encontrada em cada região. A fmp conjunta (verossimilhança) é então

$$p(x_1, \ldots, x_n; \lambda) = \frac{[\lambda \cdot a(R_1)]^{x_1} e^{-\lambda \cdot a(R_1)}}{x_1!} \cdot \cdots \cdot \frac{[\lambda \cdot a(R_n)]^{x_n} e^{-\lambda \cdot a(R_n)}}{x_n!}$$

$$= \frac{[a(R_1)]^{x_1} \cdot \cdots \cdot [a(R_n)]^{x_n} \cdot \lambda^{\Sigma x_i} \cdot e^{-\lambda \Sigma a(R_i)}}{x_1! \cdot \cdots \cdot x_n!}$$

O ln(verossimilhança) é

$$\ln[p(x_1, ..., x_n; \lambda)] = \sum x_i \cdot \ln[a(R_i)] + \ln(\lambda) \cdot \sum x_i - \lambda \sum a(R_i) - \sum \ln(x_i!)$$

Calcular $d/d\lambda \ln(p)$ ln(p) e igualá-la a zero produz

$$\frac{\sum x_i}{\lambda} - \sum a(R_i) = 0$$

então

$$\lambda = \frac{\sum x_i}{\sum a(R_i)}$$

A emv é então $\hat{\lambda} = \sum X_i / \sum a(R_i)$. Isto é intuitivamente razoável, pois λ é a densidade real (plantas por área de unidade), enquanto $\hat{\lambda}$ é a densidade da amostra, uma vez que $\sum a(R_i)$ é apenas a área total que serviu como amostra. Em virtude de $E(X_i) = \lambda \cdot a(R_i)$, o estimador é não-tendencioso.

Às vezes, é usado um procedimento alternativo de amostragem. Em vez de fixar regiões a servirem como amostra, o ecologista selecionará n pontos em toda a área de interesse e representará por y_i = a distância do i-ésimo ponto para a planta mais próxima. A função distribuição acumulada (FDA) de Y = distância para a planta mais próxima é

$$F_Y(y) = P(Y \le y) = 1 - P(Y > y) = 1 - P\left(\begin{matrix}\text{nenhuma planta em um} \\ \text{círculo de raio } y\end{matrix}\right)$$

$$= 1 - \frac{e^{-\lambda \pi y^2}(\lambda \pi y^2)^0}{0!} = 1 - e^{-\lambda \cdot \pi y^2}$$

Calculando a derivada de $F_Y(y)$ em relação a y produz

$$f_Y(y; \lambda) = \begin{cases} 2\pi\lambda y e^{-\lambda \pi y^2} & y \ge 0 \\ 0 & \text{caso contrário} \end{cases}$$

Se agora montarmos a verossimilhança $f_Y(y_1; \lambda) \cdot \cdots \cdot f_Y(y_n; \lambda)$, diferenciando ln(verossimilhança), e assim por diante, a emv resultante é

$$\hat{\lambda} = \frac{n}{\pi \sum Y_i^2} = \frac{\text{número de plantas observadas}}{\text{área total amostrada}}$$

que também é a densidade amostral. Pode-se demonstrar que, em um ambiente esparso (λ pequeno), o método da distância é, de certo modo, melhor, enquanto em um ambiente denso, o primeiro método de amostragem é melhor. ■

Exemplo 6.18

Seja $X_1, ..., X_n$ a amostra aleatória de uma fdp de Weibull

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{\alpha}{\beta^\alpha} \cdot x^{\alpha-1} \cdot e^{-(x/\beta)^\alpha} & x \ge 0 \\ 0 & \text{caso contrário} \end{cases}$$

Escrever a verossimilhança e ln(verossimilhança), depois, definir $(\partial/\partial\alpha)[\ln(f)] = 0$ e $(\partial/\partial\beta)[\ln(f)] = 0$ produz as equações

$$\alpha = \left[\frac{\sum x_i^\alpha \cdot \ln(x_i)}{\sum x_i^\alpha} - \frac{\sum \ln(x_i)}{n}\right]^{-1} \qquad \beta = \left(\frac{\sum x_i^\alpha}{n}\right)^{1/\alpha}$$

Essas duas equações não podem ser resolvidas explicitamente para fornecer fórmulas gerais para as funções $\hat{\alpha}$ e $\hat{\beta}$. Ao contrário, para cada amostra $x_1, ..., x_n$, as equações devem ser resolvidas por um procedimento numérico iterativo. Mesmo os estimadores pelo método dos momentos para α e β são um tanto complicados (veja o Exercício 21). ■

Estimando as Funções de Parâmetros

No Exemplo 6.16, obtivemos a emv de σ^2 quando a distribuição subjacente era normal. A emv de $\sigma = \sqrt{\sigma^2}$, bem como tantas outras emvs, pode ser facilmente deduzida, usando-se a seguinte proposição.

PROPOSIÇÃO **Princípio de Invariância**

Sejam $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m$ as emvs dos parâmetros $\theta_1, \theta_2, ..., \theta_m$. Então, a emv de qualquer função $h(\theta_1, \theta_2, ..., \theta_m)$ desses parâmetros é a função $h(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m)$ das emvs.

Exemplo 6.19 (continuação do Exemplo 6.16)

No caso normal, as emvs de μ e σ^2 são $\hat{\mu} = \overline{X}$ e $\hat{\sigma}^2 = \sum(X_i - \overline{X})^2/n$. Para obter a emv da função $h(\mu, \sigma^2) = \sqrt{\sigma^2} = \sigma$, substitua as emvs na função:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \left[\frac{1}{n}\sum(X_i - \overline{X})^2\right]^{1/2}$$

A emv de σ não é o desvio padrão amostral S, embora sejam próximos, a menos que n seja relativamente pequeno. ■

Exemplo 6.20 (continuação do Exemplo 6.18)

O valor médio de uma va X que tenha distribuição de Weibull é

$$\mu = \beta \cdot \Gamma(1 + 1/\alpha)$$

A emv de μ é, portanto, $\hat{\mu} = \hat{\beta}\Gamma(1 + 1/\hat{\alpha})$, onde $\hat{\alpha}$ e $\hat{\beta}$ são emvs de α e β. Em particular, $\overline{X}$ não é a emv de μ, embora seja um estimador não-tendencioso. Pelo menos para n grande, $\hat{\mu}$ é um estimador melhor que $\overline{X}$. ■

Comportamento de Amostra Grande da EMV

Embora o princípio da estimativa da máxima verossimilhança tenha apelação intuitiva considerável, a seguinte proposição fornece base adicional para o uso das emvs.

PROPOSIÇÃO

Sob condições muito genéricas, a distribuição conjunta da amostra, quando o tamanho amostral n é grande, o estimador da máxima verossimilhança de qualquer parâmetro θ é aproximadamente não-tendencioso $[E(\hat{\theta}) \approx \theta]$ e tem variância tão pequena quanto a que pode ser atingida por qualquer estimador. Dito de outra maneira, a emv $\hat{\theta}$ é aproximadamente o ENTMV de θ.

Em virtude desse resultado e do fato de as técnicas com base no cálculo geralmente poderem ser usados para deduzir as emvs (embora com freqüência sejam necessários métodos numéricos, como o método de

Newton), a estimativa de máxima verossimilhança é a técnica de estimativa mais amplamente usada entre os estatísticos. Muitos dos estimadores usados no restante do livro são emvs. Obter uma emv, entretanto, requer que a distribuição subjacente seja especificada.

Algumas Complicações

Às vezes, o cálculo não pode ser usado para obter as emvs.

Exemplo 6.21

Suponha que meu tempo de espera por um ônibus seja uniformemente distribuído em $[0, \theta]$ e que os resultados $x_1, ..., x_n$ de uma amostra aleatória dessa distribuição tenham sido observados. Uma vez que $f(x; \theta) = 1/\theta$ para $0 \le x \le \theta$ e 0, caso contrário,

$$f(x_1, ..., x_n; \theta) = \begin{cases} \dfrac{1}{\theta^n} & 0 \le x_1 \le \theta, \ldots, 0 \le x_n \le \theta \\ 0 & \text{caso contrário} \end{cases}$$

Contanto que $\max(x_i) \le \theta$, a verossimilhança é $1/\theta^n$, que é positiva, mas logo que $\theta < \max(x_i)$, a verossimilhança cai para 0. Este fato é ilustrado na Figura 6.4. O cálculo não funcionará porque o máximo da verossimilhança ocorre em um ponto de descontinuidade, mas a figura mostra que $\theta = \max(X_i)$. Assim, se meus tempos de espera são 2,3; 3,7; 1,5; 0,4 e 3,2, então a emv é $\hat{\theta} = 3,7$.

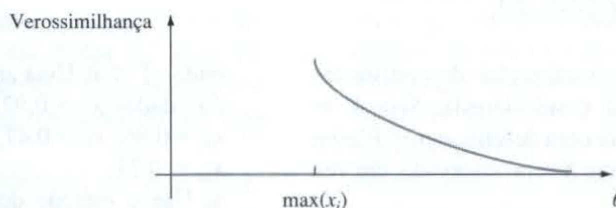


Figura 6.4 A função verossimilhança do Exemplo 6.21 ■

Exemplo 6.22

Um método geralmente usado para calcular o tamanho de uma população de animais selvagens envolve a realização de um experimento de captura/recaptura no qual uma amostra inicial de M animais é capturada, todos são identificados e devolvidos à população. Após permitir tempo suficiente para os animais marcados se misturarem na população, outra amostra de tamanho n é capturada. Com X = número de animais marcados na segunda amostra, o objetivo é usar o x observado para estimar o tamanho da população N.

O parâmetro de interesse é $\theta = N$, que pode assumir somente valores inteiros; assim, mesmo após determinar a função verossimilhança (fmp de X aqui), usar o cálculo para obter N apresentaria dificuldades. Se considerarmos um sucesso um animal anteriormente marcado que está sendo recapturado, então a amostragem é sem reposição de uma população que contém M sucessos e $N - M$ falhas, de modo que X é uma va hipergeométrica e a função verossimilhança é

$$p(x; N) = h(x; n, M, N) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

Com a natureza de N estimada em valor inteiro, entretanto, seria difícil calcular a derivada de $p(x; N)$. No entanto, se considerarmos a relação entre $p(x; N)$ a $p(x; N - 1)$, teremos

$$\frac{p(x; N)}{p(x; N-1)} = \frac{(N - M) \cdot (N - n)}{N(N - M - n + x)}$$

Essa relação é maior que 1 se e somente se $N < Mn/x$. O valor de N para o qual $p(x; N)$ é maximizado é, portanto, o maior inteiro inferior a Mn/x. Se usarmos notação matemática-padrão $[r]$ para o maior inteiro menor ou igual a r, a emv de N é $\hat{N} = [Mn/x]$. Como ilustração, se $M = 200$ peixes que são tirados de um lago e marcados, subseqüentemente, $n = 100$ peixes que são recapturados, e entre os 100 há $x = 11$ peixes marcados, então $\hat{N} = [(200)(100)/11] = [1818,18] = 1818$. A estimativa é realmente bastante intuitiva; x/n é a proporção da amostra recapturada que é marcada, enquanto M/N é a proporção da população inteira marcada. A estimativa é obtida igualando-se essas duas proporções (estimando a proporção de uma população pela proporção de uma amostra). ■

Suponha que $X_1, X_2, ..., X_n$ seja uma amostra aleatória de uma fdp $f(x; \theta)$ simétrica em relação a θ, mas que o investigador esteja incerto quanto à forma da função f. Então, é desejável usar um estimador $\hat{\theta}$ que seja robusto – isto é, que tenha bom desempenho para uma variedade ampla de fdps subjacentes. Tal estimador é uma média aparada. Nos últimos anos, os estatísticos propuseram outro tipo de estimador, chamado *Estimador M*, com base em uma generalização da estimativa de máxima verossimilhança. Em vez de maximizar o logaritmo da verossimilhança $\sum \ln[f(x; \theta)]$ para um f especificado, maximiza-se $\sum \rho(x_i; \theta)$. A "função objetivo" ρ é selecionada para produzir um estimador com boas propriedades de robustez. O livro de David Hoaglin et al. (veja a bibliografia) contém uma boa exposição sobre esse assunto.

Exercícios Seção 6.2 (20–30)

20. Uma amostra aleatória de n capacetes de ciclista fabricados por uma empresa é selecionada. Seja X = número entre os n que esteja com defeito, e $p = P$(com defeito). Assuma que somente X seja observado, em vez da seqüência de Ss e Fs.

a. Encontre o estimador de máxima verossimilhança de p. Se $n = 20$ e $x = 3$, qual é a estimativa?

b. O estimador da parte (a) é não-tendencioso?

c. Se $n = 20$ e $x = 3$, qual é a emv da probabilidade $(1 - p)^5$ de nenhum dos cinco capacetes seguintes examinados estar com defeito?

21. Seja X uma distribuição de Weibull com parâmetros α e β, então

$$E(X) = \beta \cdot \Gamma(1 + 1/\alpha)$$
$$V(X) = \beta^2\{\Gamma(1 + 2/\alpha) - [\Gamma(1 + 1/\alpha)]^2\}$$

a. Com base em uma amostra aleatória $X_1, \ldots, X_n$, escreva equações para o método dos estimadores pelos momentos de β e α. Mostre que, uma vez obtida a estimativa de α, a estimativa de β é determinada numa tabela da função gama, e que a estimativa de α é a solução para uma equação complicada que envolve a função gama.

b. Se $n = 20$, $\bar{x} = 28,0$, e $\sum x_i^2 = 16.500$, calcule as estimativas. [*Sugestão*: $[\Gamma(1,2)]^2/\Gamma(1,4) = 0,95$.]

22. Seja X a proporção de tempo distribuído que um aluno selecionado aleatoriamente gasta trabalhando em um teste de aptidão. Suponha que a fdp de X seja

$$f(x; \theta) = \begin{cases} (\theta + 1)x^\theta & 0 \le x \le 1 \\ 0 & \text{caso contrário} \end{cases}$$

onde $-1 < \theta$. Uma amostra aleatória de 10 alunos produz dados $x_1 = 0,92$, $x_2 = 0,79$, $x_3 = 0,90$, $x_4 = 0,65$, $x_5 = 0,86$, $x_6 = 0,47$, $x_7 = 0,73$, $x_8 = 0,97$, $x_9 = 0,94$, $x_{10} = 0,77$.

a. Use o método dos momentos para obter um estimador de θ e calcule a estimativa para esses dados.

b. Obtenha o estimador de máxima verossimilhança de θ e calcule a estimativa dos dados fornecidos.

23. Dois sistemas diferentes de computador são monitorados durante um total de n semanas. Seja X_i o número de falhas do primeiro sistema durante a i-ésima semana e suponha que os X_i's sejam independentes e tirados de uma distribuição de Poisson com parâmetro λ_1. De forma semelhante, seja Y_i o número de falhas do segundo sistema durante a i-ésima semana, e assuma independência em que cada Poisson de Y_i possua parâmetro λ_2. Derive as emvs de λ_1, λ_2, e $\lambda_1 - \lambda_2$. [*Sugestão*: usando a independência, escreva a fmp conjunta (verossimilhança) dos X_i's e Y_i's juntos.]

24. Consulte o Exercício 20. Em vez de selecionar $n = 20$ capacetes para examinar, suponha que eu examine os capacetes em sucessão até que encontre $r = 3$ defeituosos. Se o 20º capacete for o terceiro com defeito (de modo que o número de capacetes examinados que não tinham defeito fosse $x = 17$), qual é a emv de p? Isso é o mesmo que a estimativa do Exercício 20? Por que ou por que não? É o mesmo que a estimativa calculada pelo estimador não-tendencioso do Exercício 17?

25. A resistência de corte das 10 soldas de ponto de teste é determinada, produzindo os seguintes dados (psi):

392 376 401 367 389 362 409 415 358 375

a. Assumindo que a resistência de corte seja normalmente distribuída, estime a resistência de corte média real e o desvio padrão da resistência de corte, usando o método da máxima verossimilhança.

b. Novamente assumindo uma distribuição normal, estime o valor da resistência abaixo do qual 95% das soldas terão a resistência especificada. (*Sugestão*: qual é o 95º percentil em termos de μ e σ? Agora, use o princípio da invariância.)

26. Consulte o Exercício 25. Suponha que decidamos examinar outra solda de ponto de teste. Seja X = a resistência de corte da solda. Use os dados fornecidos para obter a emv de $P(X \leq 400)$. [*Sugestão*: $P(X \leq 400) = \Phi((400 - \mu)/\sigma)$.]

27. Seja $X_1, ..., X_n$ a amostra aleatória de uma distribuição gama com parâmetros α e β.

a. Encontre as equações cuja solução produz os estimadores de máxima verossimilhança de α e β. Você acha que eles podem ser resolvidos explicitamente?

b. Mostre que a emv de $\mu = \alpha\beta$ é $\hat{\mu} = \overline{X}$.

28. Seja $X_1, X_2, ..., X_n$ a amostra aleatória da distribuição de Rayleigh com função de densidade dada no Exercício 15. Determine

a. o estimador de máxima verossimilhança de θ e depois calcule a estimativa dos dados da força vibratória fornecida nesse exercício. Esse estimador é o mesmo que o estimador não-tendencioso sugerido no Exercício 15?

b. a emv da mediana da distribuição da força vibratória. (*Sugestão*: expresse primeiro a mediana em termos de θ.)

29. Considere uma amostra aleatória $X_1, X_2, ..., X_n$ da fdp exponencial deslocada

$$f(x; \lambda, \theta) = \begin{cases} \lambda e^{-\lambda(x-\theta)} & x \geq \theta \\ 0 & \text{caso contrário} \end{cases}$$

Considerando $\theta = 0$ fornece a fdp da distribuição exponencial considerada anteriormente (com densidade positiva à direita de zero). Um exemplo da distribuição exponencial deslocada apareceu no Exemplo 4.4, em que a variável de interesse era o tempo de avanço no fluxo do tráfego, e $\theta = 0{,}5$ era o tempo de avanço mínimo possível.

a. Obtenha os estimadores de máxima verossimilhança de θ e λ.

b. Se são feitas $n = 10$ observações do tempo de avanço, resultando nos valores 3,11, 0,64, 2,55, 2,20, 5,44, 3,42, 10,39, 8,93, 17,82 e 1,30, calcule as estimativas de θ e λ.

30. No tempo $t = 0$; 20 componentes idênticos são examinados. A distribuição da vida útil de cada um é exponencial com parâmetro λ. O investigador deixa a instalação de teste sem monitoramento. Retornando 24 horas depois, o pesquisador termina imediatamente o teste, após perceber que $y = 15$ dos 20 componentes ainda estão funcionando (então, 5 falharam). Encontre a emv de k. (*Sugestão*: seja Y = número que sobrevive 24 horas. Então, $Y \sim \text{Bin}(n, p)$. Qual é a emv de p? Agora, observe que $p = P(X_i \geq 24)$, em que X_i é exponencialmente distribuído. O que relaciona λ a p, de modo que o primeiro pode ser estimado, uma vez que o último o foi.]

Exercícios Suplementares (31–38)

31. Um estimador θ é dito ser **consistente** se para qualquer $\epsilon > 0$, $P(|\hat{\theta} - \theta| \geq \epsilon) \leftrightarrow 0$ quando $n \leftrightarrow \infty$. Isto é, $\hat{\theta}$ é consistente se, à medida que o tamanho da amostra aumenta, menor é a probabilidade de $\hat{\theta}$ ser maior que ϵ a partir do valor real de θ. Mostre que $\overline{X}$ é um estimador consistente de μ quando $\sigma^2 < \infty$ usando a desigualdade de Chebyshev do Exercício 43, do Capítulo 3. (*Sugestão*: a desigualdade pode ser reescrita na forma

$$P(|Y - \mu_Y| \geq \epsilon) \leq \sigma_Y^2/\epsilon$$

Agora, identifique Y com $\overline{X}$.)

32. a. Seja $X_1, ..., X_n$ a amostra aleatória de uma distribuição uniforme em $[0, \theta]$ Então, a emv de θ é $\hat{\theta} = Y = \max(X_i)$. Use o fato de $Y \leq y$ se e somente se cada $X_i \leq y$ para encontrar a fdc de Y. Então, mostre que a fdp de $Y = \max(X_i)$ é

$$f_Y(y) = \begin{cases} \dfrac{ny^{n-1}}{\theta^n} & 0 \leq y \leq \theta \\ 0 & \text{caso contrário} \end{cases}$$

b. Use o resultado da parte (a) para mostrar que a emv é tendenciosa, mas que $(n + 1)\max(X_i)/n$ é não-tendenciosa.

33. No tempo $t = 0$, há um indivíduo com vida em certa população. Um **processo de nascimento puro**, então, desenvolve-se como segue. O tempo até o primeiro nascimento é exponencialmente distribuído com parâmetro λ. Após o primeiro nascimento, há dois indivíduos com vida. O tempo até que o primeiro dê à luz novamente é exponencial com parâmetro λ e acontece de maneira semelhante com o segundo indivíduo. Dessa forma, o tempo até o próximo nascimento é o mínimo de duas variáveis exponenciais (λ), que é exponencial com parâmetro 2λ. De forma semelhante, uma vez que

tenha ocorrido o segundo nascimento, há três indivíduos com vida, de modo que o tempo até o nascimento seguinte é uma va exponencial com parâmetro 3λ e assim por diante (a propriedade de ausência de memória da distribuição exponencial está sendo usada aqui). Suponha que o processo seja observado até que o sexto nascimento tenha ocorrido e que os tempos de nascimento sucessivos sejam 25,2, 41,7, 51,2, 55,5, 59,5, 61,8 (a partir dos quais você deve calcular os tempos entre os nascimentos sucessivos). Encontre a emv de λ. (*Sugestão*: a verossimilhança é um produto de termos exponenciais.)

34. A **média dos erros quadráticos** de um estimador $\hat{\theta}$ é $\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$. Se $\hat{\theta}$ é não-tendencioso, então $\text{MSE}(\hat{\theta}) = V(\hat{\theta})$, mas em geral $\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + (\text{desvio})^2$. Considere o estimador $\hat{\sigma}^2 = KS^2$, em que S^2 = variância da amostra. Que valor de K minimiza a média dos erros quadráticos desse estimador quando a distribuição da população é normal? [*Sugestão*: pode-se mostrar que

$$E[(S^2)^2] = (n + 1)\sigma^4/(n - 1)$$

Em geral, é difícil achar que $\hat{\theta}$ minimiza $\text{MSE}(\hat{\theta})$, motivo pelo qual olhamos somente para estimadores não-tendenciosos e minimizamos $V(\hat{\theta})$.]

35. Seja $X_1, ..., X_n$ a amostra aleatória de uma fdp que seja simétrica com relação a μ. Um estimador de μ considerado por realizar bem uma variedade de distribuições subjacentes é o *estimador de Hodges-Lehmann*. Para defini-lo, calcule primeiro para cada $i \leq j$ e cada $j = 1, 2, ..., n$ a média de pareados $\overline{X}_{i,j} = (X_i + X_j)/2$. Então, o estimador é $\hat{\mu}$ = a mediana dos $\overline{X}_{i,j}$s. Calcule o valor dessa estimativa, usando os dados do Exercício 44, do Capítulo 1. (*Sugestão*: construa uma tabela quadrada com os x_is listados na margem esquerda e na parte superior. Depois, calcule as médias na diagonal e acima dela.)

36. Quando a distribuição da população é normal, a mediana estatística $\{\,|X_1 - \tilde{X}|, ..., |X_n - \tilde{X}|\,\}/0{,}6745$ é usada para estimar σ. O estimador é mais resistente aos efeitos de *outliers* (observações distantes da parte principal dos dados) que o desvio padrão da amostra. Calcule a estimativa pontual correspondente e s para os dados do Exemplo 6.2.

37. Quando o desvio padrão da amostra S se baseia em uma amostra aleatória de uma distribuição normal da população, pode ser demonstrado que

$$E(S) = \sqrt{2/(n-1)}\,\Gamma(n/2)\sigma/\Gamma((n-1)/2)$$

Use essa equação para obter um estimador não-tendencioso para σ da forma cS. Qual é c quando $n = 20$?

38. Todos os n espécimes serão pesados duas vezes na mesma escala. Represente por X_i e Y_i os dois pesos observados do i-ésimo espécime. Suponha que X_i e Y_i sejam independentes entre si, cada um normalmente distribuído com valor médio μ_i (o peso real do espécime i) e variância σ^2.

a. Mostre que o estimador de máxima verossimilhança de σ^2 é $\hat{\sigma}^2 = \sum(X_i - Y_i)^2/(4n)$. [*Sugestão*: se $\bar{z} = (z_1 + z_2)/2$, então, $\sum(z_i - \bar{z})^2 = (z_1 - z_2)^2/2$.]

b. A emv $\hat{\sigma}^2$ é um estimador não-tendencioso de σ^2? Determine um estimador não-tendencioso de σ^2. [*Sugestão*: para qualquer va Z, $E(Z^2) = V(Z) + [E(Z)]^2$. Aplique a equação em $Z = X_i - Y_i$.]

Bibliografia

DEGROOT, Morris e SCHERVISH, Mark. *Probability and Statistics* (3. ed.). Addison-Wesley: Boston, MA, 2002. Inclui uma excelente discussão das propriedades e métodos gerais da estimativa pontual; para interesse particular, existem exemplos que mostram como os princípios e métodos gerais podem produzir estimadores insatisfatórios em situações específicas.

EFRON, Bradley e TIBSHIRANI, Robert. *An Introduction to the Bootstrap,* Chapman e Hall, Nova York, 1993. A bíblia do *bootstrap*.

HOAGLIN, David, MOSTELLER, Frederick e TUKEY, John. *Understanding Robust and Exploratory Data Analysis*. Wiley, Nova York, 1983. Contém diversos capítulos bons sobre estimativa pontual robusta, incluindo um sobre Estimativa *M*.

HOGG, Robert e CRAIG, Allen. *Introduction to Mathematical Statistics*. 5. ed., Prentice Hall, Englewood Cliffs, NJ, 1995. Uma boa discussão sobre não-tendenciosidade.

LARSEN, Richard e MARX, Morris. *Introduction to Mathematical Statistics*. 2. ed., Prentice Hall, Englewood Cliffs, NJ, 1985. Uma discussão muito boa sobre estimativa pontual a partir de uma perspectiva levemente mais matemática que o texto atual.

RICE, John. *Mathematical Statistics and Data Analysis*. 2. ed., Duxbury Press, Belmont, CA, 1994. Uma boa mistura de teoria estatística e dados.

7

Intervalos Estatísticos Baseados em uma Única Amostra

Introdução

Uma estimativa pontual, por ser um único número, não fornece por si mesma qualquer informação sobre a precisão e a confiabilidade da estimativa. Considere, por exemplo, o uso da estatística $\overline{X}$ para calcular a estimativa pontual da tensão de quebra média real (g) de toalhas de papel de certa marca e suponha que $\overline{x} = 9322,7$. Devido à variabilidade da amostragem, virtualmente nunca é o caso de $\overline{x} = \mu$. A estimativa pontual não diz nada sobre o quanto pode estar próxima de μ. Uma alternativa para apresentar um único valor sensato para o parâmetro que está sendo estimado é calcular e relatar um intervalo completo de valores plausíveis — uma *estimativa de intervalo* ou *intervalo de confiança* (IC). Um intervalo de confiança sempre é calculado selecionando-se primeiro o *nível de confiança*, que é uma medida do grau de confiabilidade do intervalo. Um intervalo de confiança com nível de confiança de 95% da tensão de quebra média real pode ter limite inferior de 9162,5 e limite superior de 9482,9. Então, com nível de confiança de 95%, qualquer valor de μ entre 9162,5 e 9482,9 é plausível. O nível de confiança de 95% implica que 95% de todas as amostras forneceriam um intervalo que inclui μ, ou qualquer outro parâmetro que esteja sendo estimado, e apenas 5% de todas as amostras dariam um intervalo errôneo. Os níveis de confiança usados com mais freqüência são 95%, 99% e 90%. Quanto maior o nível de confiança, mais fortemente acreditamos que o valor do parâmetro que está sendo estimado está dentro do intervalo (a interpretação de qualquer nível de confiança específico será dada em breve).

As informações sobre a precisão de uma estimativa de intervalo são transmitidas pela sua extensão. Se o nível de confiança for alto e o intervalo resultante, bastante restrito, nosso conhecimento do valor do parâmetro será razoavelmente preciso. Um intervalo de confiança muito amplo, entretanto, passa a idéia de que há muita incerteza com relação ao valor do que estamos estimando. A Figura 7.1 mostra intervalos de confiança de 95% das tensões de quebra médias reais de duas marcas diferentes de toalhas de papel. Um

desses intervalos sugere conhecimento exato de μ, enquanto o outro sugere uma variedade bastante grande de valores plausíveis.

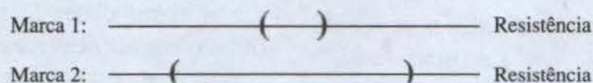


Figura 7.1 Intervalos de confiança que indicam informações precisas (marca 1) e imprecisas (marca 2) sobre μ

7.1 Propriedades Básicas de Intervalos de Confiança

Os conceitos básicos e as propriedades dos intervalos de confiança (ICs) são introduzidos mais facilmente enfocando primeiro uma situação de problema simples, embora um pouco irreal. Suponha que o parâmetro de interesse seja a média de uma população μ e que

1. A distribuição da população é normal
2. O valor do desvio padrão da população σ é conhecido

A normalidade da distribuição populacional é geralmente uma suposição razoável. Entretanto, se o valor de μ é desconhecido, é improvável que o valor de σ esteja disponível (o conhecimento do centro de uma população precede tipicamente informações relacionadas à dispersão). Nas seções posteriores, desenvolveremos métodos com base em suposições menos restritivas.

Exemplo 7.1

Os engenheiros industriais que se especializam em ergonomia estão preocupados em projetar espaços e dispositivos operados por trabalhadores, de modo a obter maior produtividade e conforto. O artigo "Studies on Ergonomically Designed Alphanumeric Keyboards" (*Human Factors,* 1985, p. 175-187) relata o estudo de altura preferida de um teclado experimental com grande apoio para o pulso e o antebraço. Uma amostra de $n = 31$ digitadores treinados foi selecionada, e a altura preferida do teclado foi determinada para cada digitador. A altura preferida média resultante da amostra foi $\bar{x} = 80$ cm. Assumindo que a altura preferida seja normalmente distribuída com $\sigma = 2$ cm (valor sugerido pelos dados no artigo), obtenha um IC para μ, a altura média real preferida pela população de todos os digitadores experientes. ▮

As observações reais da amostra $x_1, x_2, ..., x_n$ são consideradas o resultado de uma amostra aleatória $X_1, ..., X_n$ de uma distribuição normal com valor médio μ e desvio padrão σ. Os resultados do Capítulo 5, então, implicam que, independentemente do tamanho da amostra n, a média amostral $\bar{X}$ é normalmente distribuída com valor esperado μ e desvio padrão $\sigma/\sqrt{n}$. Padronizar $\bar{X}$ subtraindo primeiro seu valor esperado e dividindo-o por seu desvio padrão produz a variável

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (7.1)$$

que tem distribuição normal padronizada. Em virtude de a área sob a curva normal padrão entre −1,96 e 1,96 ser 0,95,

$$P\left(-1{,}96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1{,}96\right) = 0{,}95 \quad (7.2)$$

O passo seguinte no desenvolvimento é manipular as desigualdades dentro dos parênteses em (7.2), para que apareçam sob a forma equivalente $l < \mu < u$, em que os pontos finais l e u envolvam $\overline{X}$ e $\sigma/\sqrt{n}$. Isso é alcançado por meio da seguinte seqüência de operações, cada uma produzindo desigualdades equivalentes às que iniciamos com:

1. Multiplique por $\sigma/\sqrt{n}$ para obter

$$-1{,}96 \cdot \frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}$$

2. Subtraia $\overline{X}$ de cada termo para obter

$$-\overline{X} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}$$

3. Multiplique por -1 para eliminar o sinal negativo de μ (que inverte a direção de cada desigualdade) para obter

$$\overline{X} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} > \mu > \overline{X} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}$$

isto é,

$$\overline{X} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}$$

Em virtude de cada conjunto de desigualdades da seqüência ser equivalente ao original, a probabilidade associada a cada um é 0,95. Em particular

$$P\left(\overline{X} - 1{,}96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 0{,}95 \qquad (7.3)$$

O evento dentro dos parênteses em (7.3) possui uma aparência um tanto desconhecida; antes a quantidade aleatória sempre apareceu no meio com constantes em ambas as extremidades, como em $a \leq Y \leq b$. Em (7.3), a quantidade aleatória aparece nas duas extremidades, enquanto a constante desconhecida μ aparece no meio. Para interpretar (7.3), imagine um **intervalo aleatório** que tem ponto final esquerdo $\overline{X} - 1{,}96 \cdot \sigma/\sqrt{n}$ e ponto final direito $\overline{X} + 1{,}96 \cdot \sigma/\sqrt{n}$ que, na notação de intervalo, é

$$\left(\overline{X} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}\right) \qquad (7.4)$$

O intervalo (7.4) é aleatório porque seus dois pontos finais envolvem uma variável aleatória (va). Observe que o intervalo está centrado na média amostral $\overline{X}$ e se estende $1{,}96\sigma/\sqrt{n}$ de cada lado de $\overline{X}$. Dessa forma, a amplitude do intervalo é $2 \cdot (1{,}96) \cdot \sigma/\sqrt{n}$, que não é aleatório; somente a posição do intervalo (seu ponto central $\overline{X}$) é aleatória (Figura 7.2). Agora, (7.3) pode ser parafraseado como *"a probabilidade de que o intervalo aleatório* (7.4) *inclua o valor real de μ é* 0,95*"*. Antes de qualquer experimento ser realizado e de quaisquer dados serem coletados, é bastante provável (probabilidade de 0,95) que μ esteja dentro do intervalo na Expressão (7.4).

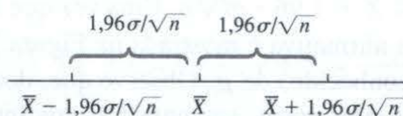


Figura 7.2 O intervalo aleatório (7.4) centralizado em $\overline{X}$

DEFINIÇÃO

Após observar $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$, calculamos a média amostral observada $\bar{x}$ e substituímos $\bar{x}$ em (7.4) no lugar de $\bar{X}$; o intervalo fixo resultante é chamado **intervalo de confiança de 95% de μ**. Esse IC pode ser expresso tanto como

$$\left(\bar{x} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}\right) \text{ é um IC de 95\% de } \mu$$

ou como

$$\bar{x} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} \quad \text{com 95\% de confiança}$$

Uma expressão concisa para o intervalo é $\bar{x} \pm 1{,}96 \cdot \sigma/\sqrt{n}$, onde $-$ é o ponto final esquerdo (limite inferior) e $+$ é o ponto final direito (limite superior).

Exemplo 7.2 (continuação do Exemplo 7.1)

As quantidades requeridas para o cálculo do IC de 95% da altura média real preferida são $\sigma = 2{,}0$, $n = 31$ e $\bar{x} = 80{,}0$. O intervalo resultante é

$$\bar{x} \pm 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} = 80{,}0 \pm (1{,}96)\frac{2{,}0}{\sqrt{31}} = 80{,}0 \pm 0{,}7 = (79{,}3, 80{,}7)$$

Isto é, podemos estar altamente confiantes, no nível de confiança de 95%, de que $79{,}3 < \mu < 80{,}7$. Esse intervalo é relativamente restrito, indicando que μ foi estimado com precisão. ■

Interpretando um Intervalo de Confiança

O nível de confiança de 95% do intervalo definido foi herdado da probabilidade de 0,95 do intervalo aleatório (7.4). Os intervalos que têm outros níveis de confiança serão introduzidos em breve. Agora, no entanto, considere como a confiança de 95% pode ser interpretada.

Em virtude de termos começado com um evento cuja probabilidade era 0,95 – que o intervalo aleatório (7.4) incluiria o valor real de μ – e depois usado os dados do Exemplo 7.1 para calcular o intervalo estabelecido (79,3, 80,7), é tentador concluir que μ está dentro desse intervalo estabelecido com probabilidade de 0,95. Mas, substituindo $\bar{x} = 80{,}0$ por $\bar{X}$, toda a aleatoriedade desaparece; o intervalo (79,3, 80,7) não é aleatório e μ é uma constante (infelizmente desconhecida para nós), de modo que é *errado* escrever a expressão $P(\mu$ está em (79,3, 80,7)) = 0,95.

A interpretação correta da "confiança de 95%" baseia-se na interpretação de freqüência relativa de longo prazo da probabilidade: dizer que um evento A tem probabilidade de 0,95 significa dizer que, se o experimento no qual A é definido for repetido várias vezes, em longo prazo, A ocorrerá 95% das vezes. Suponha que obtenhamos outra amostra das alturas preferidas dos digitadores e calculemos outro intervalo de 95%. Então, consideramos repetir o experimento para uma terceira amostra, para uma quarta e assim por diante. Seja A o evento em que $\bar{X} - 1{,}96 \cdot \sigma/\sqrt{n} < \mu < \bar{X} + 1{,}96 \cdot \sigma/\sqrt{n}$. Uma vez que $P(A) = 0{,}95$, em longo prazo, 95% de nossos ICs calculados conterão μ. Esta afirmativa é mostrada na Figura 7.3, onde a linha vertical corta o eixo de medição no valor real (embora desconhecido) de μ. Observe que, dos 11 intervalos ilustrados, somente os intervalos 3 e 11 não contêm μ. Em longo prazo, somente 5% dos intervalos construídos desse modo não contêm μ.

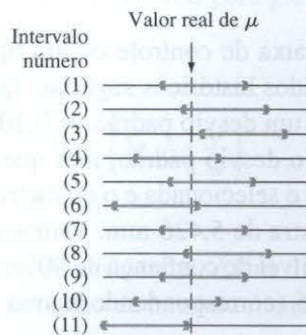


Figura 7.3 Construção repetida dos ICs de 95%

Conforme essa interpretação, o nível de confiança de 95% não é tanto uma declaração sobre qualquer intervalo específico, como (79,3, 80,7), mas se refere ao que aconteceria se um número muito grande de intervalos semelhantes fosse construído. Embora isso possa parecer insatisfatório, a origem da dificuldade permanece com nossa interpretação de probabilidade; aplica-se a uma seqüência longa de replicações de um experimento em vez de apenas uma única replicação. Existe outra abordagem para a construção e interpretação de ICs que utiliza a noção de probabilidade subjetiva e o teorema de Bayes, mas os detalhes técnicos estão além do escopo deste texto; o livro de Winkler (veja a bibliografia do Capítulo 2) é uma boa fonte. O intervalo apresentado aqui (bem como cada intervalo apresentado subseqüentemente) é chamado IC "clássico", pois sua interpretação depende da noção clássica de probabilidade (embora as idéias principais tenham sido desenvolvidas em um passado relativamente recente, na década de 1930).

Outros Níveis de Confiança

O nível de confiança de 95% foi herdado da probabilidade de 0,95 das desigualdades iniciais em (7.2). Se for desejado um nível de confiança de 99%, a probabilidade inicial de 0,95 deve ser substituída por 0,99, o que exige a mudança do valor crítico de z de 1,96 para 2,58. Um IC de 99% resulta, então, do uso de 2,58 no lugar de 1,96 na fórmula do IC de 95%.

Essa afirmativa sugere que qualquer nível de confiança desejado pode ser alcançado substituindo-se 1,96 ou 2,58 pelo valor crítico normal padronizado apropriado. Conforme mostra a Figura 7.4, a probabilidade de $1 - \alpha$ é alcançada usando-se $z_{\alpha/2}$ no lugar de 1,96.

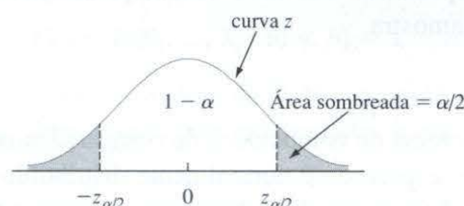


Figura 7.4 $P(-z_{\alpha/2} \leq Z < z_{\alpha/2}) = 1 - \alpha$

DEFINIÇÃO

O intervalo de confiança 100(1 − α)% da média μ de uma população normal, quando o valor de σ é conhecido, é dado por

$$\left(\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \tag{7.5}$$

ou, de forma equivalente, por $\overline{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

Exemplo 7.3

O processo de produção das unidades de caixa de controle de um tipo específico de motor foi modificado recentemente. Antes dessa modificação, os dados históricos sugeriam que a distribuição dos diâmetros do orifício dos mancais nas caixas eram normais, com um desvio padrão de 0,100 mm. Acredita-se que a modificação não tenha afetado o formato da distribuição ou o desvio padrão, mas que o valor do diâmetro médio possa ter mudado. Uma amostra de 40 unidades da caixa é selecionada e o diâmetro do orifício é determinado para cada uma, resultando em um diâmetro médio da amostra de 5,426 mm. Vamos calcular um intervalo de confiança para o diâmetro médio real do orifício usando um nível de confiança de 90%. O procedimento requer que $100(1-\alpha) = 90$, do qual $\alpha = 0,10$ e $z_{\alpha/2} = z_{0,05} = 1,645$ (correspondendo a uma área acumulada da curva z de 0,9500). O intervalo desejado é, então,

$$5,426 \pm (1,645)\frac{0,100}{\sqrt{40}} = 5,426 \pm 0,026 = (5,400, 5,452)$$

Com um grau razoavelmente alto de confiança, podemos dizer que $5,400 < \mu < 5,452$. Esse intervalo é bastante restrito, devido ao pequeno valor da variabilidade do diâmetro do orifício ($\sigma = 0,100$). ∎

Nível de Confiança, Precisão e Escolha do Tamanho da Amostra

Por que seguir para um nível de confiança de 95% quando se pode chegar a um nível de 99%? Porque o preço pago pelo nível de confiança maior é um intervalo mais largo. Em virtude de o intervalo de 95% estender-se $1,96 \cdot \sigma/\sqrt{n}$ a cada lado de $\bar{x}$, a amplitude do intervalo é $2(1,96) \cdot \sigma/\sqrt{n} = 3,92 \cdot \sigma/\sqrt{n}$. De maneira semelhante, a amplitude do intervalo de 99% é $2(2,58) \cdot \sigma/\sqrt{n} = 5,16 \cdot \sigma/\sqrt{n}$. Isto é, temos mais confiança no intervalo de 99% exatamente porque é mais largo. Assim, quanto maior o grau de confiança desejado, mais largo é o intervalo resultante. Na verdade, o único IC de 100% para μ é $(-\infty, \infty)$, que não acrescenta grandes informações, pois, mesmo antes da amostragem, sabíamos que esse intervalo cobria μ.

Se imaginarmos a amplitude do intervalo como a especificação de sua precisão ou acurácia, então seu nível de confiança (ou confiabilidade) estará inversamente relacionado à sua precisão. Uma estimativa altamente confiável do intervalo pode ser imprecisa, quando os pontos finais do intervalo estiverem muito distantes, enquanto um intervalo preciso pode exigir confiabilidade relativamente baixa. Dessa forma, não se pode dizer inequivocamente que um intervalo de 99% será preferível a um intervalo de 95%; o ganho na confiabilidade exige uma perda na precisão.

Uma estratégia interessante é especificar o nível de confiança desejado e a amplitude do intervalo e então determinar o tamanho necessário da amostra.

Exemplo 7.4

O monitoramento extensivo de um sistema de computador de compartilhamento de tempo sugeriu que o tempo de resposta a um comando de edição específico é normalmente distribuído com desvio padrão de 25 milissegundos. Um novo sistema operacional foi instalado e desejamos estimar o tempo de resposta médio real μ do novo ambiente. Assumindo que os tempos de resposta ainda sejam normalmente distribuídos com $\sigma = 25$, que tamanho de amostra é necessário para garantir que o IC de 95% resultante tenha uma amplitude de (no máximo) 10? O tamanho da amostra n deve satisfazer

$$10 = 2 \cdot (1,96)(25/\sqrt{n})$$

O reajuste dessa equação fornece

$$\sqrt{n} = 2 \cdot (1,96)(25)/10 = 9,80$$

então

$$n = (9,80)^2 = 96,04$$

Uma vez que n deve ser um número inteiro, é necessário um tamanho de amostra de 97. ∎

A fórmula geral do tamanho da amostra n necessária para garantir uma amplitude de intervalo w é obtida de $w = 2 \cdot z_{\alpha/2} \cdot \sigma/\sqrt{n}$, já que

$$n = \left(2z_{\alpha/2} \cdot \frac{\sigma}{w}\right)^2$$

Quanto menor a amplitude desejada w, maior deve ser n. Além disso, n é uma função crescente de σ (maior variabilidade da população precisa de um tamanho de amostra maior) e do nível de confiança $100(1 - \alpha)$ (à medida que α diminui, $z_{\alpha/2}$ aumenta).

A metade da amplitude $1{,}96\sigma/\sqrt{n}$ do IC de 95% é chamada às vezes de **limite do erro de estimação** associado a um nível de confiança de 95%; isto é, com confiança de 95%, a estimativa pontual $\bar{x}$ não estará mais distante de μ do que este valor. Antes de obter os dados, o investigador pode desejar determinar o tamanho de uma amostra para o qual é obtido um valor específico do limite. Por exemplo: com μ representando o consumo médio de combustível (milhas/galão) de todos os carros de um determinado tipo, o objetivo de uma investigação pode ser estimar μ dentro de um intervalo de confiança de 95% de 1 milha/galão. Geralmente, se desejamos estimar μ para um valor B (o limite especificado do erro de estimação) com confiança de $100(1-\alpha)\%$, o tamanho da amostra necessário resulta da substituição de $2/w$ por $1/B$ na fórmula da caixa precedente.

Derivando um Intervalo de Confiança

Sejam $X_1, X_2, ..., X_n$, a amostra na qual o IC de um parâmetro θ se baseará. Suponha que uma variável aleatória que satisfaça às duas propriedades a seguir possa ser encontrada:

1. A variável depende funcionalmente de $X_1, ..., X_n$ e θ.
2. A distribuição de probabilidade da variável não depende de θ ou de quaisquer outros parâmetros desconhecidos.

Seja $h(X_1, X_2, ..., X_n; \theta)$ essa variável aleatória. Por exemplo: se a distribuição da população for normal com σ conhecido e $\theta = \mu$, a variável $h(X_1, ..., X_n; \mu) = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ satisfaz ambas as propriedades; sua forma funcional claramente depende de μ, ainda que tenha distribuição de probabilidade normal padronizada, que não depende de μ. Em geral, a forma da função h é sugerida examinando-se a distribuição de um estimador apropriado θ.

Para qualquer α entre 0 e 1, podem ser determinadas constantes a e b que satisfaçam a

$$P(a < h(X_1, ..., X_n; \theta) < b) = 1 - \alpha \tag{7.6}$$

Devido à segunda propriedade, a e b não dependem de θ. No exemplo normal, $a = -z_{\alpha/2}$ e $b = z_{\alpha/2}$. Agora, suponha que as desigualdades em (7.6) possam ser manipuladas para isolar θ, fornecendo a probabilidade equivalente

$$P(l(X_1, X_2, ..., X_n) < \theta < u(X_1, X_2, ..., X_n)) = 1 - \alpha$$

Então, $l(x_1, x_2, ..., x_n)$ e $u(x_1, ..., x_n)$ são os limites de confiança inferior e superior, respectivamente, de um IC de $100(1 - \alpha)\%$. No exemplo normal, vimos que $l(X_1, ..., X_n) = \bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n}$ e $u(X_1, ..., X_n) = \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

Exemplo 7.5

Um modelo teórico sugere que o tempo de quebra de um líquido isolante entre eletrodos em uma voltagem específica tem distribuição exponencial com parâmetro λ (veja a Seção 4.4). Uma amostra aleatória de $n = 10$ tempos de quebra produz os seguintes dados amostrais (em min): $x_1 = 41{,}53$, $x_2 = 18{,}73$, $x_3 = 2{,}99$, $x_4 = 30{,}34$, $x_5 = 12{,}33$, $x_6 = 117{,}52$, $x_7 = 73{,}02$, $x_8 = 223{,}63$, $x_9 = 4{,}00$, $x_{10} = 26{,}78$. Um IC de 95% de λ e do tempo de quebra médio real são desejados.

Seja $h(X_1, X_2, ..., X_n; \lambda) = 2\lambda\sum X_i$. Demonstra-se que essa variável aleatória possui uma distribuição de probabilidade chamada distribuição qui-quadrada com $2n$ graus de liberdade (gl) ($\nu = 2n$, onde ν é o parâmetro da distribuição qui-quadrada, conforme discutido na Seção 4.4). A Tabela A.7 do Apêndice ilustra uma curva de densidade qui-quadrada típica e tabula valores críticos que incluem as áreas de cauda especificadas. O número relevante de graus de liberdade neste caso é $2(10) = 20$. A linha $\nu = 20$ da tabela mostra que 34,170 inclui a área sob a cauda superior 0,025 e 9,591 inclui a área sob a cauda inferior 0,025 (área sob a cauda superior 0,975). Assim, para $n = 10$,

$$P(9{,}591 < 2\lambda\sum X_i < 34{,}170) = 0{,}95$$

Divisão por $2\sum X_i$ isola λ, produzindo

$$P(9{,}591/(2\sum X_i) < \lambda < (34{,}170/(2\sum X_i)) = 0{,}95$$

O limite inferior do IC de 95% de λ é $9{,}591/(2\sum x_i)$, e o limite superior é $34{,}170/(2\sum x_i)$. Para os dados fornecidos, $\sum x_i = 550{,}87$, dando o intervalo (0,00871, 0,03101).

O valor esperado de uma va exponencial é $\mu = 1/\lambda$. Uma vez que

$$P(2\sum X_i/34{,}170 < 1/\lambda < 2\sum X_i/9{,}591) = 0{,}95$$

o IC de 95% do tempo de quebra médio real é $(2\sum x_i/34{,}170,\ 2\sum x_i/9{,}591) = (32{,}24,\ 114{,}87)$. Esse intervalo é evidentemente bastante largo, refletindo variabilidade substancial nos tempos de quebra e o tamanho de uma amostra pequena. ■

Em geral, os limites de confiança superior e inferior resultam da substituição de cada < em (7.6) por = e da resolução de θ. No exemplo do fluido isolante considerado, $2\lambda\sum x_i = 34{,}170$ fornece $\lambda = 34{,}170/(2\sum x_i)$ como o limite de confiança superior e o limite inferior, obtidos da outra equação. Observe que os dois limites do intervalo não são eqüidistantes da estimativa pontual, uma vez que o intervalo não é da forma $\hat{\theta} \pm c$.

Intervalos de Confiança *Bootstrap*

A técnica *bootstrap* foi introduzida no Capítulo 6 como uma forma de estimar $\sigma_{\hat{\theta}}$. Também pode ser aplicada para se obter um IC de θ. Considere novamente a estimação da média μ de uma distribuição normal quando σ é conhecido. Vamos substituir μ por θ e usar $\hat{\theta} = \overline{X}$ como o estimador pontual. Observe que $1{,}96\sigma/\sqrt{n}$ é o 97,5º percentil da distribuição de $\hat{\theta} - \theta$ [isto é, $P(\overline{X} - \mu < 1{,}96\sigma/\sqrt{n}) = P(Z < 1{,}96) = 0{,}9750$]. De forma semelhante, $-1{,}96\sigma/\sqrt{n}$ é o 2,5º percentil, assim

$$\begin{aligned} 0{,}95 &= P(2{,}5^{\underline{o}} \text{ percentil} < \hat{\theta} - \theta < 97{,}5^{\underline{o}} \text{ percentil}) \\ &= P(\hat{\theta} - 2{,}5^{\underline{o}} \text{ percentil} > \theta > \hat{\theta} - 97{,}5^{\underline{o}} \text{ percentil}) \end{aligned}$$

Isto é, com

$$\begin{aligned} l &= \hat{\theta} - 97{,}5^{\underline{o}} \text{ percentil de } \hat{\theta} - \theta \\ u &= \hat{\theta} - 2{,}5^{\underline{o}} \text{ percentil de } \hat{\theta} - \theta \end{aligned} \tag{7.7}$$

o IC de θ é (l, u). Em muitos casos, os percentis em (7.7) não podem ser calculados, mas *podem* ser estimados a partir das amostras *bootstrap*. Suponha que obtenhamos $B = 1000$ amostras *bootstrap* e calcule $\hat{\theta}_1^*, ..., \hat{\theta}_{1000}^*$ e $\overline{\theta}^*$ e, depois, as 1000 diferenças $\hat{\theta}_1^* - \overline{\theta}^*, ..., \hat{\theta}_{1000}^* - \overline{\theta}^*$. A 25ª maior e a 25ª menor dessas diferenças são estimativas dos percentis desconhecidos em (7.7). Consulte os livros de Rice ou Efron, citados no Capítulo 6, para mais informações.

Exercícios Seção 7.1 (1–11)

1. Considere a distribuição de uma população normal com o valor de σ conhecido.
   a. Qual é o nível de confiança do intervalo $\bar{x} \pm 2{,}81\sigma/\sqrt{n}$?
   b. Qual é o nível de confiança do intervalo $\bar{x} \pm 1{,}44\sigma/\sqrt{n}$?
   c. Que valor de $z_{\alpha/2}$ na fórmula do IC (7.5) resulta em um nível de confiança de 99,7%?
   d. Responda à pergunta proposta no item (c) para um nível de confiança de 75%.

2. Cada um dos intervalos a seguir é um intervalo de confiança de μ = média real (isso é, média da população) da freqüência de ressonância (Hz) de todas as raquetes de tênis de um determinado tipo:

   (114,4, 115,6)   (114,1, 115,9)

   a. Qual é o valor da freqüência de ressonância da média amostral?
   b. Ambos os intervalos foram calculados a partir dos mesmos dados amostrais. O nível de confiança de um desses intervalos é 90% e do outro é 99%. Qual dos intervalos possui o nível de confiança de 90% e por quê?

3. Suponha que uma amostra aleatória de 50 garrafas de uma marca específica de xarope para tosse seja selecionada e o teor alcoólico de cada garrafa seja determinado. Seja μ o teor médio de álcool da população de todas as garrafas da marca em estudo. Suponha que o intervalo de confiança de 95% resultante seja (7,8, 9,4).
   a. Um intervalo de confiança de 90% calculado dessa mesma amostra teria sido mais estreito ou mais largo que o intervalo mencionado acima? Explique seu raciocínio.
   b. Considere a afirmação a seguir: existe 95% de chance de μ estar entre 7,8 e 9,4. Essa afirmação está correta? Por quê?
   c. Considere a afirmação a seguir: podemos estar certos de que 95% de todas as garrafas desse tipo de xarope têm um conteúdo alcoólico que está entre 7,8 e 9,4. A afirmação está correta? Por quê?
   d. Considere a afirmação a seguir: se o processo de seleção de uma amostra de tamanho 50 e de cálculo do intervalo de 95% correspondente for repetido 100 vezes, 95 dos intervalos resultantes incluirão μ. Essa afirmação está correta? Por quê?

4. Deseja-se um IC para a média real da perda de carga por dispersão μ (watts) de um tipo de motor a indução, quando a corrente da linha é mantida em 10 amps para uma velocidade de 1500 rpm. Assuma que a perda de carga por dispersão seja normalmente distribuída com $\sigma = 3{,}0$.
   a. Calcule um IC de 95% de μ quando $n = 25$ e $\bar{x} = 58{,}3$.
   b. Calcule um IC de 95% de μ quando $n = 100$ e $\bar{x} = 58{,}3$.
   c. Calcule um IC de 99% de μ quando $n = 100$ e $\bar{x} = 58{,}3$.
   d. Calcule um IC de 82% de μ quando $n = 100$ e $\bar{x} = 58{,}3$.
   e. Quão grande deve ser n se a largura do intervalo de 99% de μ for 1,0?

5. Assuma que a porosidade do hélio (em porcentagem) das amostras de carvão tiradas de qualquer junta específica seja normalmente distribuída com desvio padrão real de 0,75.
   a. Calcule um IC de 95% da porosidade média real de uma junta, caso a porosidade média de 20 de seus espécimes seja 4,85.
   b. Calcule um IC de 98% da porosidade média real de outra junta com base nos 16 espécimes com média amostral de porosidade de 4,56.
   c. Quão grande o tamanho de uma amostra deve ser se a amplitude do intervalo de 95% for 0,40?
   d. Que tamanho de amostra é necessário para estimar a porosidade média real dentro de 0,2 com confiança de 99%?

6. Com base em testes extensivos, o ponto de escoamento de um tipo específico de barra de aço reforçado é conhecido por ser normalmente distribuído com $\sigma = 100$. A composição da barra foi levemente modificada, mas acredita-se que essa modificação não tenha afetado nem a normalidade e nem o valor de σ.
   a. Assumindo que esse seja o caso, se uma amostra de 25 barras modificadas tiver resultado em um ponto de escoamento médio amostral de 8439 lb, calcule um IC de 90% para o ponto de escoamento médio real da barra modificada.
   b. Como você modificaria o intervalo no item (a) para obter um nível de confiança de 92%?

7. Em quanto o tamanho da amostra n deve ser aumentado se a amplitude do IC (7.5) for reduzida à metade? Se o tamanho da amostra for aumentado por um fator de 25, que efeito terá sobre a amplitude do intervalo? Justifique suas afirmações.

8. Seja $\alpha_1 > 0$, $\alpha_2 > 0$, com $\alpha_1 + \alpha_2 = \alpha$. Então,

$$P\left(-z_{\alpha_1} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha_2}\right) = 1 - \alpha$$

   a. Use essa equação para deduzir uma expressão mais geral para um IC de $100(1 - \alpha)\%$ de μ do qual o intervalo (7.5) é um caso especial.
   b. Seja $\alpha = 0{,}05$ e $\alpha_1 = \alpha/4$, $\alpha_2 = 3\alpha/4$. Essa condição resulta em um intervalo mais estreito ou mais largo que o intervalo (7.5)?

9. a. Sob as mesmas condições que as que levam ao intervalo (7.5), $P[(\overline{X} - \mu)/(\sigma/\sqrt{n}) < 1{,}645] = 0{,}95$. Use isso para deduzir um intervalo unilateral de μ que possui amplitude infinita e fornece um limite de confiança inferior em μ. Que intervalo é esse para os dados do Exercício 5(a)?

b. Generalize o resultado do item (a) para obter um limite inferior com nível de confiança $100(1 - \alpha)\%$.

c. Qual é o intervalo análogo ao do item (b) que fornece um limite superior em μ? Calcule esse intervalo de 99% para os dados do Exercício 4(a).

10. Uma amostra aleatória de $n = 15$ determinados tipos de bombas de aquecimento produziu as seguintes observações sobre a vida útil (em anos):

2,0	1,3	6,0	1,9	5,1	0,4	1,0	5,3
15,7	0,7	4,8	0,9	12,2	5,3	0,6	

a. Assuma que a distribuição da vida útil seja exponencial e use um argumento paralelo ao do Exemplo 7.5 para obter um IC de 95% da vida útil esperada (médio real).

b. Como o intervalo do item (a) deve ser alterado para chegar a um nível de confiança de 99%?

c. O que é o IC de 95% para o desvio padrão da distribuição da vida útil? (*Sugestão*: qual é o desvio padrão de uma variável aleatória exponencial?)

11. Considere os próximos 1000 ICs de 95% de μ que um consultor estatístico obterá para vários clientes. Suponha que os conjuntos de dados em que os intervalos se baseiam sejam selecionados independentemente um do outro. Quantos desses 1000 intervalos capturam os valores correspondentes de μ? Qual é a probabilidade, de que 940 e 960 desses intervalos contenham o valor correspondente de μ? (*Sugestão*: seja Y = número de intervalos entre os 1000 que contém μ. Que tipo de variável aleatória é Y?)

7.2 Intervalos de Confiança para Amostras Grandes para uma Média e Proporção da População

O IC de μ dado na seção anterior assumiu que a distribuição da população é normal e que o valor de σ é conhecido. Apresentamos agora o IC de uma amostra grande, cuja validade não exige essas hipóteses. Depois de mostrar como o argumento que leva a esse intervalo é generalizado para produzir outros de amostra grande, abordamos o intervalo da proporção de uma população p.

Um Intervalo de Amostra Grande para μ

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória de uma população que tem média μ e desvio padrão σ. Desde que n seja grande, o Teorema do Limite Central (TLC) implica que $\overline{X}$ tem aproximadamente distribuição normal, qualquer que seja a natureza da distribuição da população. Segue então que $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ possui aproximadamente distribuição normal padrão, de modo que

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \approx 1 - \alpha$$

Um argumento paralelo ao fornecido na Seção 7.1 produz $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$ como IC de amostra grande de μ com nível de confiança de *aproximadamente* $100(1 - \alpha)\%$. Isto é, quando n é grande, o IC de μ fornecido anteriormente permanece válido, qualquer que seja a distribuição da população, contanto que o adjetivo "aproximadamente" se insira na frente do nível de confiança.

Uma dificuldade prática com esse desenvolvimento é que o cálculo do intervalo exige o valor de σ, que quase nunca será conhecido. Considere a variável padronizada

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

em que o desvio padrão amostral S substitui σ. Havia anteriormente aleatoriedade somente no numerador de Z (em virtude de $\overline{X}$). Agora, existe aleatoriedade no numerador e no denominador – os valores de $\overline{X}$ e S variam de amostra para amostra. Entretanto, quando n é grande, o uso de S em vez de σ acrescenta pouca variabilidade extra a Z. Mais especificamente, nesse caso, o novo Z também possui aproximadamente uma distribuição normal padronizada. A manipulação das desigualdades para probabilidades que utilizam esse novo Z produz um intervalo geral de amostras grandes para μ.

PROPOSIÇÃO

Se n é suficientemente grande, a variável padronizada

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

possui aproximadamente uma distribuição normal padronizada. Isso implica que

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \qquad (7.8)$$

é um **intervalo de confiança de amostra grande para μ** com nível de confiança de aproximadamente $100(1 - \alpha)\%$. Essa fórmula é válida independentemente do formato da distribuição da população.

Geralmente, $n > 40$ será suficiente para justificar o uso desse intervalo. Essa afirmação é mais conservadora que a regra prática para o TLC, devido à variabilidade adicional introduzida, usando-se S em vez de σ.

Exemplo 7.6

A voltagem de quebra da Corrente Alternada (CA) de um líquido isolante indica sua resistência dielétrica. O artigo "Testing Practices for the AC Breakdown Voltage Testing of Insulation Liquids" (*IEEE Electrical Insulation Magazine*, 1995, p. 21-26) forneceu as seguintes observações amostrais da voltagem de quebra (kV) de um circuito específico sob certas condições.

62	50	53	57	41	53	55	61	59	64	50	53	64	62	50	68
54	55	57	50	55	50	56	55	46	55	53	54	52	47	47	55
57	48	63	57	57	55	53	59	53	52	50	55	60	50	56	58

Um *boxplot* dos dados (Figura 7.5) mostra alta concentração central dos dados (largura estreita da caixa). Há um único *outlier* na extremidade superior, mas esse valor está realmente um pouco mais próximo da mediana (55) que a observação da amostra menor.

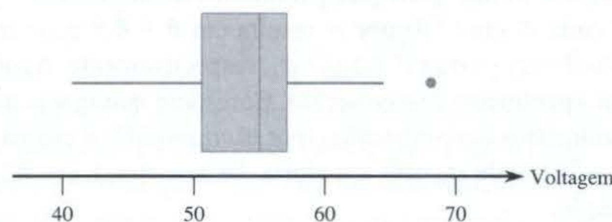


Figura 7.5 Um *boxplot* para os dados da voltagem de quebra do Exemplo 7.6

As quantidades resumidas incluem $n = 48$, $\sum x_i = 2626$ e $\sum x_i^2 = 144{,}950$, pelas quais $\bar{x} = 54{,}7$ e $s = 5{,}23$. O intervalo de confiança de 95% é então

$$54{,}7 \pm 1{,}96\frac{5{,}23}{\sqrt{48}} = 54{,}7 \pm 1{,}5 = (53{,}2,\ 56{,}2)$$

Isto é,

$$53{,}2 < \mu < 56{,}2$$

com um nível de confiança de aproximadamente 95%. O intervalo é razoavelmente estreito, indicando que estimamos μ com precisão. ■

Infelizmente, a escolha do tamanho da amostra para produzir a largura desejada de um intervalo não é tão direta aqui como foi para o caso de sigma conhecido. Isso se deve ao fato de a largura de (7.8) ser $2z_{\alpha/2}s/\sqrt{n}$. Uma vez que o valor de s não está disponível antes da coleta dos dados, a amplitude do intervalo não pode ser determinada somente pela escolha de n. A única opção para um investigador que quer especificar uma amplitude desejada é fazer uma estimativa sensata sobre qual pode ser o valor de s. Sendo conservador e estimando um valor maior de s, será escolhido um n maior do que o necessário. O investigador pode ser capaz de especificar um valor razoavelmente preciso para a amplitude da população (a diferença entre os valores menor e maior). Então, se a distribuição da população não estiver muito distorcida, dividir a amplitude por 4 proporciona um valor aproximado de s.

Exemplo 7.7

Consulte o Exemplo 7.6 sobre a voltagem de quebra. Suponha que o investigador acredite que, virtualmente, todos os valores da população estejam entre 40 e 70. Então, $(70 - 40)/4 = 7{,}5$ é um valor razoável de s. O tamanho apropriado da amostra para estimar a média real da voltagem de quebra para dentro de 1 kV com nível de confiança de 95% — isto é, para o IC de 95% ter amplitude de 2 kV — é

$$n = [(1{,}96)(7{,}5)/1]^2 \approx 217$$

■

Um Intervalo de Confiança Geral para Amostra Grande

Os intervalos de amostra grande $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$ e $\bar{x} \pm z_{\alpha/2} \cdot s/\sqrt{n}$ são casos especiais de IC geral para amostra grande para um parâmetro θ. Suponha que $\hat{\theta}$ seja um estimador que satisfaça às seguintes propriedades: (1) Possui distribuição aproximadamente normal; (2) é (pelo menos aproximadamente) não-tendencioso; e (3) está disponível uma expressão de $\sigma_{\hat{\theta}}$, o desvio padrão de $\hat{\theta}$. Por exemplo: no caso $\theta = \mu$, $\hat{\mu} = \bar{X}$ é um estimador não-tendencioso cuja distribuição é aproximadamente normal quando n é grande e $\sigma_{\hat{\mu}} = \sigma_{\bar{X}} = \sigma/\sqrt{n}$. Padronizar $\hat{\theta}$ produz a va $Z = (\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$ que possui distribuição aproximadamente normal padronizada. Isso justifica a declaração de probabilidade.

$$P\left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} < z_{\alpha/2}\right) \approx 1 - \alpha \tag{7.9}$$

Suponha, primeiro, que $\sigma_{\hat{\theta}}$ não utilize quaisquer parâmetros desconhecidos (por exemplo: σ conhecido no caso $\theta = \mu$). Então, substituir cada $<$ em (7.9) por $=$ resulta em $\theta = \hat{\theta} \pm z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$, de modo que os limites de confiança inferior e superior são $\hat{\theta} - z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$ e $\hat{\theta} + z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$, respectivamente. Agora, suponha que $\sigma_{\hat{\theta}}$ não utilize θ, mas que use pelo menos outro parâmetro desconhecido. Considere que $s_{\hat{\theta}}$ seja a estimativa de $\sigma_{\hat{\theta}}$ obtida usando-se estimativas em vez de parâmetros desconhecidos (por exemplo: $s/\sqrt{n}$ estima $\sigma/\sqrt{n}$). Sob condições gerais (essencialmente que $s_{\hat{\theta}}$ esteja próximo de $\sigma_{\hat{\theta}}$ para a maioria das amostras), um IC válido é $\hat{\theta} \pm z_{\alpha/2} \cdot s_{\hat{\theta}}$. O intervalo $\bar{x} \pm z_{\alpha/2} \cdot s/\sqrt{n}$ é um exemplo.

Finalmente, suponha que $\sigma_{\hat{\theta}}$ utilize o θ desconhecido. Esse é o caso, por exemplo, quando $\theta = p$, uma proporção da população. Então, $(\hat{\theta} - \theta)/\sigma_{\hat{\theta}} = z_{\alpha/2}$ pode ser difícil de resolver. Uma solução aproximada geralmente pode ser obtida substituindo-se θ em $\sigma_{\hat{\theta}}$ por sua estimativa $\hat{\theta}$. Resulta, pois, um desvio padrão estimado $s_{\hat{\theta}}$ e o intervalo correspondente é novamente $\theta \pm z_{\alpha/2} \cdot s_{\hat{\theta}}$.

Um Intervalo de Confiança para uma Proporção da População

Seja p a proporção de "sucessos" de uma população em que *sucesso* identifica um indivíduo ou objeto que tenha uma propriedade especificada. Uma amostra aleatória de n indivíduos será selecionada e X é o número de sucessos na amostra. Contanto que n seja pequeno, em comparação ao tamanho da população, X pode ser considerado uma va binomial com $E(X) = np$ e $\sigma_X = \sqrt{np(1 - p)}$. Além disso, se n for grande ($np \geq 10$ e $nq \geq 10$), X possui distribuição aproximadamente normal.

O estimador natural de p é $\hat{p} = X/n$, a fração de sucessos da amostra. Uma vez que $\hat{p}$ é apenas X multiplicado pela constante $1/n$, $\hat{p}$ também tem distribuição aproximadamente normal. Como mostrado na Seção 6.1,

$E(\hat{p}) = p$ (não-tendencioso) e $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. O desvio padrão $\sigma_{\hat{p}}$ utiliza o parâmetro desconhecido p. Padronizar $\hat{p}$, pela subtração de p e divisão por $\sigma_{\hat{p}}$, implica, então, que

$$P\left(-z_{\alpha/2} < \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} < z_{\alpha/2}\right) \approx 1-\alpha$$

Prosseguindo conforme foi sugerido na subseção "Derivando um Intervalo de Confiança" (Seção 7.1), os limites de confiança resultam da substituição de cada < por = e resolvendo a equação quadrática resultante de p. Obtêm-se as duas raízes

$$p = \frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z^2_{\alpha/2}}{4n^2}}}{1 + (z^2_{\alpha/2})/n}$$

PROPOSIÇÃO

O **intervalo de confiança para uma proporção da população** p com nível de confiança de aproximadamente $100(1-\alpha)\%$ tem

$$\text{limite de confiança inferior} = \frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z^2_{\alpha/2}}{4n^2}}}{1 + (z^2_{\alpha/2})/n}$$

e (7.10)

$$\text{limite de confiança superior} = \frac{\hat{p} + \frac{z^2_{\alpha/2}}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z^2_{\alpha/2}}{4n^2}}}{1 + (z^2_{\alpha/2})/n}$$

Se o tamanho da amostra for muito grande, $z^2/(2n)$ é desprezível, comparado a $\hat{p}$, $z^2/(4n^2)$ sob a raiz quadrada é desprezível se comparado a $\hat{p}\hat{q}/n$, e z^2/n é desprezível se comparado a 1. Desconsiderando esses termos desprezíveis obtemos os limites de confiança aproximados

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} \qquad (7.11)$$

Essa expressão tem a forma geral $\hat{\theta} \pm z_{\alpha/2}\,\hat{\sigma}_{\hat{\theta}}$ de um intervalo de amostra grande sugerido na última subseção. Durante décadas, o intervalo acima foi recomendado, desde que a aproximação normal de $\hat{p}$ fosse justificada. Entretanto uma pesquisa recente mostrou que o intervalo um tanto mais complicado dado na proposição tem um nível de confiança real que tende a estar mais próximo do nível nominal do que o intervalo tradicional (Agresti, Alan e Coull, "Approximate Is Better Than 'Exact' for Interval Estimation of a Binomial Proportion," *The American Statistician*, 1998, p. 119-126). Isto é, se $z_{\alpha/2} = 1{,}96$ for usado, o nível de confiança do "novo" intervalo tende a estar mais próximo de 95% para quase todos os valores de p que é o caso do intervalo tradicional; isto também acontece com outros níveis de confiança. Além disso, Agresti e Coull afirmam que o intervalo "pode ser recomendado para uso em aproximadamente todos os tamanhos de amostra e valores de parâmetro", de modo que as condições $n\hat{p} \geq 10$ e $n\hat{q} \geq 10$ não precisem ser verificadas.

Exemplo 7.8

O artigo "Repeatability and Reproducibility for Pass/Fail Data" (*J. of Testing and Eval.*, 1997, p. 151-153) relatou que em $n = 48$ tentativas em um laboratório específico, 16 resultaram em ignição de um tipo específico de substrato por um cigarro aceso. Seja p a proporção no longo prazo de todas as tentativas que resultariam em ignição. A estimativa pontual de p é $\hat{p} = 16/48 = 0{,}333$. O intervalo de confiança de p com nível de confiança de aproximadamente 95% é

$$\frac{0{,}333 + (1{,}96)^2/96 \pm 1{,}96\sqrt{(0{,}333)(0{,}667)/48 + (1{,}96)^2/9216}}{1 + (1{,}96)^2/48} = \frac{0{,}373 \pm 0{,}139}{1{,}08} = (0{,}217, 0{,}474)$$

O intervalo tradicional é

$$0{,}333 \pm 1{,}96\sqrt{(0{,}333)(0{,}667)/48} = 0{,}333 \pm 1{,}33 = (0{,}200, 0{,}466)$$

Esses dois intervalos estariam muito mais próximos entre si se o tamanho da amostra fosse substancialmente maior. ∎

Igualar a amplitude do IC de p a uma amplitude pré-especificada w fornece uma equação quadrática para o tamanho da amostra n necessário para fornecer um intervalo com grau de precisão desejado. Eliminando-se o índice de $z_{\alpha/2}$, a solução é

$$n = \frac{2z^2\hat{p}\hat{q} - z^2w^2 \pm \sqrt{4z^4\hat{p}\hat{q}(\hat{p}\hat{q} - w^2) + w^2z^4}}{w^2} \quad (7.12)$$

Desprezando os termos do numerador que envolvem w^2, temos

$$n \approx \frac{4z^2\hat{p}\hat{q}}{w^2}$$

Essa última expressão é o resultado da igualdade da amplitude do intervalo tradicional com w.

Essas fórmulas infelizmente envolvem $\hat{p}$ desconhecido. A abordagem mais conservadora é tirar proveito do fato de $\hat{p}\hat{q}$ [$= \hat{p}(1 - \hat{p})$] ser um máximo quando $\hat{p} = 0{,}5$. Assim, se $\hat{p} = \hat{q} = 0{,}5$ for usado em (7.12), a amplitude será no máximo w, independentemente do valor de $\hat{p}$ que resulte da amostra. De maneira alternativa, se o investigador acredita fortemente, com base em informações anteriores, que $p \leq p_0 \leq 0{,}5$, então p_0 pode ser usado no lugar de $\hat{p}$. Um comentário semelhante se aplica quando $p \geq p_0 \geq 0{,}5$.

Exemplo 7.9

A amplitude do IC de 95% do Exemplo 7.8 é 0,257. O valor de n necessário para garantir uma amplitude de 0,10 sem considerar o valor de $\hat{p}$ é

$$n = \frac{2(1{,}96)^2(0{,}25) - (1{,}96)^2(0{,}01) \pm \sqrt{4(1{,}96)^4(0{,}25)(0{,}25 - 0{,}01) + (0{,}01)(1{,}96)^4}}{0{,}01} = 380{,}3$$

Dessa forma, deve ser usado um tamanho de amostra de 381. A expressão de n com base no IC tradicional fornece um valor levemente maior de 385. ∎

Intervalos de Confiança Monocaudais (Limites de Confiança)

Os intervalos de confiança discutidos até agora fornecem tanto um limite de confiança inferior como um limite de confiança superior para o parâmetro que está sendo estimado. Em algumas circunstâncias, o investigador vai querer somente um desses dois tipos de limites. Por exemplo: um psicólogo pode querer calcular o limite de confiança superior de 95% para o tempo de reação médio real a um estímulo específico, ou um engenheiro de confiabilidade pode querer somente o limite de confiança inferior da vida útil média real de um determinado tipo de componente. Em virtude de a área acumulada sob a curva normal padrão à esquerda de 1,645 ser 0,95,

$$P\left(\frac{\overline{X} - \mu}{S/\sqrt{n}} < 1{,}645\right) \approx 0{,}95$$

Manipular a desigualdade dentro dos parênteses para isolar μ em um lado e substituir as vas pelos valores calculados fornece a desigualdade $\mu > \bar{x} - 1{,}645s/\sqrt{n}$; a expressão à direita é o limite de confiança inferior desejado. Começar com $P(-1{,}645 < Z) \approx 0{,}95$ e manipular a desigualdade resulta no limite de confiança superior. Um argumento semelhante fornece o limite monocaudal associado a qualquer outro nível de confiança.

PROPOSIÇÃO

O limite de confiança superior de amostra grande para μ é

$$\mu < \bar{x} + z_\alpha \cdot \frac{s}{\sqrt{n}}$$

e o limite de confiança inferior de amostra grande para μ é

$$\mu > \bar{x} - z_\alpha \cdot \frac{s}{\sqrt{n}}$$

O limite de confiança monocaudal para p resulta da substituição de $z_{\alpha/2}$ por z_α e $\pm$ por + ou – na fórmula do IC (7.10) para p. Em todos os casos, o nível de confiança é de aproximadamente $100(1 - \alpha)\%$.

Exemplo 7.10

O teste de corte inclinado é o procedimento mais amplamente aceito para avaliar a qualidade de uma ligação entre um material de conserto e seu substrato de concreto. O artigo "Testing the Bond Between Repair Materials and Concrete Substrate" (*ACI Materials J.*, 1996, p. 553-558) relatou que, em uma investigação específica, uma amostra de 48 observações de resistência de corte forneceu uma resistência média amostral de 17,17 N/mm^2 e um desvio padrão da amostra de 3,28 N/mm^2. O limite de confiança inferior para a resistência de corte média real μ com nível de confiança de 95% é

$$17{,}17 - (1{,}645)\frac{(3{,}28)}{\sqrt{48}} = 17{,}17 - 0{,}78 = 16{,}39$$

Isto é, com um nível de confiança de 95%, o valor de μ está no intervalo $(16{,}39, \infty)$. ■

Exercícios | Seção 7.2 (12-27)

12. Uma amostra aleatória de 110 relâmpagos, em certa região, resultou em uma duração média do eco do radar da amostra de 0,81 s. e um desvio padrão da amostra de 0,34 s. ("Lightning Strikes to an Airplane in a Thunderstorm," *J. of Aircraft*, 1984, p. 607-611). Calcule o intervalo de confiança (bicaudal) de 99% para a duração média real do eco μ, e interprete o intervalo resultante.

13. O artigo "Extravisual Damage Detection? Defining the Standard Normal Tree" (*Photogrammetric Engr. and Remote Sensing*, 1981, p. 515-522) discute o uso de fotografia infravermelha colorida na identificação de árvores normais nos grupos de pinheiros Douglas. Dentre os dados relatados, estava a estatística resumida das medições densitométricas ópticas analíticas de filtro verde nas amostras de árvores doentes e saudáveis. Para uma amostra de 69 árvores saudáveis, a densidade da camada colorida média amostral era 1,028, e o desvio padrão da amostra, 0,163.

   a. Calcule um IC (bicaudal) de 95% para a densidade da camada colorida média real de todas as árvores.

   b. Suponha que os investigadores tenham feito uma adivinhação grosseira de 0,16 para o valor de s antes de coletar os dados. Que tamanho de amostra seria necessário para obter uma amplitude do intervalo de 0,05, para um nível de confiança de 95%?

14. O artigo "Evaluating Tunnel Kiln Performance" (*Amer. Ceramic Soc. Bull.*, ago. 1997, p. 59-63) forneceu as seguintes informações resumidas das resistências a fraturas (Mpa) de $n = 169$ barras cerâmicas cozidas em um determinado forno: $\bar{x} = 89{,}10$, $s = 3{,}73$.

a. Calcule o intervalo de confiança (bicaudal) para a resistência à fratura média real, usando um nível de confiança de 95%. Parece que a resistência à fratura média real foi estimada com precisão?

b. Suponha que os investigadores acreditaram *a priori* que o desvio padrão da população era cerca de 4 MPa. Com base nessa suposição, quão grande deve ser uma amostra para calcular μ dentro de 0,5 MPa com confiança de 95%?

15. Determine o nível de confiança para cada um dos limites de confiança monocaudais de amostras grandes a seguir:

a. Limite superior: $\bar{x} + 0{,}84s/\sqrt{n}$

b. Limite inferior: $\bar{x} - 2{,}05s/\sqrt{n}$

c. Limite superior: $\bar{x} + 0{,}67s/\sqrt{n}$

16. O tempo entre a carga e o final do processo (min) de um aço carbono em um tipo de fornalha aberta foi determinado para cada aquecimento em uma amostra de tamanho 46, resultando um tempo médio amostral de 382,1 e um desvio padrão da amostra de 31,5. Calcule o limite de confiança superior de 95% do tempo médio real entre a carga e o final do processo.

17. O Exercício 1.13 forneceu uma amostra de observações da resistência extrema à tensão (ksi). Use a saída da estatística descritiva do MINITAB a seguir para calcular o limite de confiança inferior de 99% para a resistência extrema à tensão média real e interpretar o resultado.

```
N      Média   Mediana   Média Ap  Desv Padrão  Média SE
153    135,39  135,40    135,41    4,59         0,37
Mínimo  Máximo  Q1        Q3
122,20  147,70  132,95    138,25
```

18. O artigo "Ultimate Load Capacities of Expansion Anchor Bolts" (*J. of Energy Engr.*, 1993, p. 139-158) forneceu os seguintes dados resumidos sobre a resistência de corte (kip) de uma amostra de pinos de ferro de 3/8-pol.: $n = 78$, $\bar{x} = 4{,}25$, $s = 1{,}30$. Calcule o limite de confiança inferior, usando um nível de confiança de 90% para a resistência de corte média real.

19. O artigo "Limited Yield Estimation for Visual Defect Sources" (*IEEE Trans. on Semiconductor Manuf.*, 1997, p. 17-23) relatou que, em um estudo do processo de inspeção de um *wafer* específico, 356 dados foram examinados por uma inspeção e 201 deles foram aprovados. Assumindo um processo estável, calcule o intervalo de confiança (bicaudal) de 95% para a proporção de todos os dados que foram aprovados pela inspeção.

20. A Associated Press (9 de outubro de 2002) relatou que em uma inspeção de 4722 jovens americanos, com idade entre 6 e 19 anos, 15% estavam seriamente acima do peso (índice de massa corporal de pelo menos 30; esse índice é uma medida do peso em relação à altura). Calcule e interprete um intervalo de confiança, usando um nível de confiança de 99% para a proporção de todos os jovens americanos que estão seriamente acima do peso.

21. Uma amostra aleatória de 539 lares de certa cidade do Meio-Oeste foi selecionada e determinou-se que em 133 deles havia pelo menos uma arma de fogo ("The Social Determinants of Gun Ownership: Self-Protection in an Urban Environment", *Criminology*, 1997, p. 629-640). Usando um nível de confiança de 95%, calcule o limite de confiança inferior para a proporção de todos os lares dessa cidade que possuam no mínimo uma arma de fogo.

22. Uma amostra aleatória de 487 mulheres não-fumantes de peso normal (índice de massa corporal entre 19,8 e 26,0) que deram à luz em um grande centro médico metropolitano foi selecionada ("The Effects of Cigarette Smoking and Gestational Weight Change on Birth Outcomes in Obese and Normal-Weight Women", *Amer. J. of Public Health*, 1997, p. 591-596). Determinou-se que 7,2% desses partos resultaram em crianças com baixo peso (menos de 2500 g). Calcule o limite de confiança superior, usando um nível de confiança de 99% para a proporção de todos os nascimentos que resultaram em crianças com baixo peso.

23. O artigo "An Evaluation of Football Helmets Under Impact Conditions" (*Amer. J. Sports Medicine*, 1984, p. 233-237) relata que, quando cada capacete de futebol americano, em uma amostra aleatória de 37 capacetes do tipo suspensão, foi submetida a certo teste de impacto, 24 mostraram estar com defeito. Seja p a proporção de todos os capacetes desse tipo que mostrariam algum dano quando testados da maneira prescrita.

a. Calcule o IC de 99% para p.

b. Que tamanho de amostra seria necessário para a amplitude de um IC de 99% ser no máximo de 0,10 sem considerar $\hat{p}$?

24. Uma de 56 amostras de algodão de pesquisa resultou em uma porcentagem de alongamento médio da amostra de 8,17 e desvio padrão amostral de 1,42 ("An Apparent Relation Between the Spiral Angle ϕ, the Percent Elongation E_1, and the Dimensions of the Cotton Fiber", *Textile Research J.*, 1978, p. 407-410). Calcule o IC de amostra grande de 95% para a porcentagem de alongamento médio real μ. Que suposições você está fazendo com relação à distribuição da porcentagem de alongamento?

25. Um deputado estadual deseja pesquisar os residentes de seu distrito para ver que proporção do eleitorado está ciente de sua posição sobre o uso de fundos do Estado para pagar abortos.

a. Que tamanho de amostra é necessário, se o IC de 95% de p tiver amplitude de no máximo 0,10 sem considerar p?

b. Se o deputado tem forte razão para acreditar que pelo menos $\frac{2}{3}$ do eleitorado sabe de sua posição, quão grande você recomendaria que o tamanho da amostra fosse?

26. O superintendente do distrito de uma grande escola, que fez um curso de probabilidade e estatística, acredita

que o número de professores ausentes em um dia qualquer possua distribuição de Poisson com parâmetro λ. Use os dados a seguir sobre as ausências durante 50 dias para deduzir o IC de amostra grande para λ. [*Sugestão*: a média e a variância de uma variável de Poisson são ambos iguais a λ, assim

$$Z = \frac{\overline{X} - \lambda}{\sqrt{\lambda/n}}$$

possui aproximadamente distribuição normal padronizada. Prossiga como na derivação do intervalo de p, definindo a situação (com probabilidade $1 - \alpha$) e resolvendo as desigualdades resultantes de λ (veja a explicação abaixo de (7.10)).]

Número de ausentes	0	1	2	3	4	5	6	7	8	9	10
Freqüência	1	4	8	10	8	7	5	3	2	1	1

27. Reconsidere o IC (7.10) para p e enfoque um nível de confiança de 95%. Mostre que os limites de confiança estão de acordo com os do intervalo tradicional (7.11), uma vez que dois sucessos e duas falhas foram adicionados à amostra [isto é, (7.11) com base em $x + 2$ S's em $n + 4$ tentativas]. (*Sugestão*: $1{,}96 \approx 2$. *Nota*: Agresti e Coull mostraram que esse ajuste do intervalo tradicional também possui nível de confiança real próximo ao nível nominal.)

7.3 Intervalos Baseados em uma Distribuição Normal da População

O IC de μ apresentado na Seção 7.2 é válido, desde que n seja grande. O intervalo resultante é usado, qualquer que seja a natureza da distribuição da população. O TLC não poderá ser usado, entretanto, quando n for pequeno. Nesse caso, uma maneira de continuar é fazer uma suposição específica sobre a forma da distribuição da população e deduzir um IC feito sob medida a essa suposição. Por exemplo: poderíamos desenvolver um IC de μ quando a população for descrita por uma distribuição gama, outro intervalo para o caso de uma população de Weibull, e assim por diante. Os estatísticos realmente fizeram esse programa para um número de famílias de distribuição diferentes. Em virtude de a distribuição normal ser apropriada mais freqüentemente como um modelo da população que qualquer outro tipo de distribuição, enfocaremos aqui um IC para essa situação.

SUPOSIÇÃO — A população de interesse é normal, de modo que $X_1, ..., X_n$ constitui uma amostra aleatória de uma distribuição normal com μ e σ desconhecidos.

O resultado-chave subordinado ao intervalo na Seção 7.2 foi que, para n grande, a va $Z = (\overline{X} - \mu)/(S/\sqrt{n})$ possui aproximadamente uma distribuição normal padronizada. Quando n é pequeno, S provavelmente não está mais próximo de σ, de modo que a variabilidade na distribuição de Z surge da aleatoriedade no numerador e no denominador. Isso implica que a distribuição de probabilidade de $(\overline{X} - \mu)/(S/\sqrt{n})$ será mais dispersa que a distribuição normal padronizada. O resultado no qual as inferências se baseiam introduz uma nova família de distribuições de probabilidade chamada família de *distribuições t*.

TEOREMA — Quando $\overline{X}$ é a média amostral aleatória de tamanho n de uma distribuição normal com média μ, a va

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{7.13}$$

possui uma distribuição de probabilidade chamada distribuição t com $n - 1$ graus de liberdade (gl).

Propriedades das Distribuições *t*

Antes de aplicar este teorema, uma discussão das propriedades das distribuições t é solicitada. Embora a variável de interesse ainda seja $(\overline{X} - \mu)/(S/\sqrt{n})$, nós a representamos agora por T, para enfatizar que ela não possui uma distribuição normal padronizada quando n é pequeno. Lembre-se de que uma distribuição normal é determinada por dois parâmetros, a média μ e o desvio padrão σ. A distribuição t é determinada apenas por um parâmetro, chamado **número de graus de liberdade** da distribuição, abreviado por gl. Nós o representamos pela letra grega ν. Valores possíveis de ν são os inteiros positivos 1, 2, 3, ... Cada valor diferente de ν corresponde a uma distribuição t diferente.

Para qualquer valor fixo do parâmetro ν, a função densidade que especifica a curva t associada possui uma aparência ainda mais complicada que a função densidade normal. Felizmente, precisamos nos preocupar somente com algumas das características mais importantes dessas curvas.

> **Propriedades das distribuições *t***
>
> Seja t_ν a curva da função densidade dos gl de ν.
>
> **1.** Cada curva t_ν possui formato de sino e está centrada em 0.
>
> **2.** Toda curva t_ν é mais dispersa que a curva normal padronizada (z).
>
> **3.** À medida que ν aumenta, a dispersão da curva t_ν correspondente diminui.
>
> **4.** À medida que $\nu \to \infty$ a seqüência das curvas t_ν se aproxima da curva normal padronizada (de modo que a curva z é chamada geralmente de curva t com gl $= \infty$).

A Figura 7.6 ilustra algumas dessas propriedades dos valores selecionados de ν.

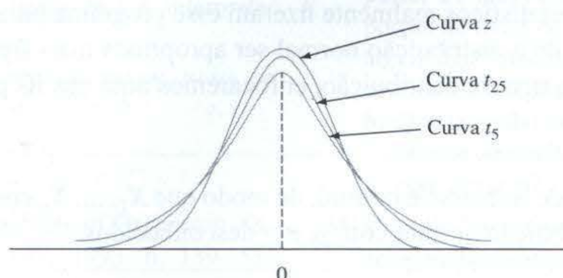


Figura 7.6 Curvas t_ν e z

O número de gl de T em (7.13) é $n - 1$, pois, embora S se baseie nos n desvios $X_1 - \overline{X}, ..., X_n - \overline{X}$, $\Sigma(X_i - \overline{X}) = 0$ implica que somente $n - 1$ deles são "livremente determinados". A quantidade de gl de uma variável t é a quantidade de desvios determinados livremente onde o desvio padrão estimado no denominador de T se baseia.

Uma vez que queremos usar T para obter um IC da mesma maneira que Z foi usado anteriormente, é necessário estabelecer notação análoga a z_α para a distribuição t.

> **Notações**
>
> Seja $t_{\alpha,\nu}$ = o número no eixo de medição para o qual a área sob a curva t com gl ν à direita de $t_{\alpha,\nu}$ é α; $t_{\alpha,\nu}$ é chamado **valor crítico *t***.

Essa notação é ilustrada na Figura 7.7. A Tabela A.5 do Apêndice fornece $t_{\alpha,\nu}$ para valores selecionados de α e ν. Essa tabela também aparece na contracapa. As colunas da tabela correspondem a valores diferentes de α. Para obter $t_{0,05,15}$, vá para a coluna $\alpha = 0,05$, procure a linha $\nu = 15$ e leia $t_{0,05,\,15} = 1,753$. De maneira semelhante, $t_{0,05,\,22} = 1,717$ (coluna 0,05, linha $\nu = 22$) e $t_{0,01,\,22} = 2,508$.

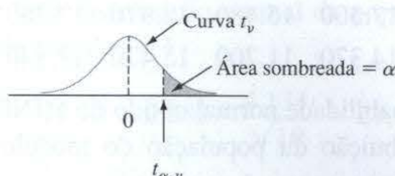


Figura 7.7 Definição ilustrada de $t_{\alpha,\nu}$

Os valores de $t_{\alpha,\nu}$ exibem comportamento regular, à medida que mudamos de linha ou coluna. Para ν fixo, $t_{\alpha,\nu}$ aumenta à medida que α diminui, uma vez que temos de mover à direita de zero para encontrar a área α na cauda. Para α fixo à medida que ν é aumentado (isto é, quando procuramos em qualquer coluna específica da tabela t), o valor de $t_{\alpha,\nu}$ diminui, porque um valor maior de ν implica uma distribuição t com dispersão menor, de modo que não é necessário se afastar muito de zero para encontrar a área α da cauda. Além disso, $t_{\alpha,\nu}$ diminui mais lentamente à medida que ν aumenta. Conseqüentemente, os valores da tabela são mostrados em incrementos de 2 entre 30 e 40 gl e depois pulam para $\nu = 50, 60, 120$ e finalmente ∞. Em virtude de t_∞ ser a curva normal padronizada, os valores z_α familiares aparecem na última linha da tabela. A regra prática sugerida anteriormente para uso do IC de amostra grande (se $n > 40$) surge da igualdade aproximada das distribuições t e normal padronizada de $\nu \geq 40$.

O Intervalo de Confiança *t* de uma Amostra

A variável padronizada T possui distribuição t com $n - 1$ gl e a área sob a curva de densidade t correspondente entre $-t_{\alpha/2,n-1}$ e $t_{\alpha/2,n-1}$ é $1 - \alpha$ (a área $\alpha/2$ está em cada cauda), assim

$$P(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}) = 1 - \alpha \tag{7.14}$$

A expressão (7.14) difere-se das expressões nas seções anteriores porque T e $t_{\alpha/2,n-1}$ são usados no lugar de Z e $z_{\alpha/2}$ mas pode ser manipulada da mesma forma para obter um intervalo de confiança de μ.

PROPOSIÇÃO

Sejam $\bar{x}$ e s a média e o desvio padrão amostrais calculados a partir dos resultados de uma amostra aleatória de uma população normal com média μ. Então, o **intervalo de confiança de 100(1 − α)% para μ** é

$$\left(\bar{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}\right) \tag{7.15}$$

ou, mais compactamente, $\bar{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n}$.

O limite de confiança superior de μ é

$$\bar{x} + t_{\alpha,n-1} \cdot \frac{s}{\sqrt{n}}$$

e substituir + por – nessa última expressão fornece o **limite de confiança inferior de μ**, com nível de confiança de $100(1 - \alpha)\%$.

Exemplo 7.11

Como parte de um projeto maior para estudar o comportamento de painéis de revestimento tencionado, componente estrutural que está sendo usado extensivamente nos Estados Unidos, o artigo "Time-Dependent Bending Properties of Lumber" (*J. of Testing and Eval.*, 1996, p. 187-193) relatou diversas propriedades mecânicas de espécimes de madeira serrada de pinho da Escócia. Considere as seguintes observações sobre o módulo de elasticidade (MPa) obtido 1 minuto depois da aplicação de carga em uma determinada configuração:

10.490	16.620	17.300	15.480	12.970	17.260	13.400	13.900
13.630	13.260	14.370	11.700	15.470	17.840	14.070	14.760

A Figura 7.8 mostra um gráfico de probabilidade normal obtido do MINITAB. O padrão reto no gráfico fornece forte apoio para assumir que a distribuição da população do módulo de elasticidade é pelo menos aproximadamente normal.

O cálculo manual da média amostral e do desvio padrão é simplificado subtraindo 10.000 de cada observação: $y_i = x_i - 10.000$. Verifica-se com facilidade que $\sum y_i = 72.520$ e $\sum y_i^2 = 392.083.800$, pelo que $\bar{y} = 4532,5$ e $s_y = 2055,67$. Dessa forma, $\bar{x} = 14.532,5$ e $s_x = 2055,67$ (adicionar ou subtrair a mesma constante de cada observação não afeta a variabilidade). O tamanho da amostra é 16, de modo que o intervalo de confiança do módulo de elasticidade médio da população se baseia em 15 gl. Um nível de confiança de 95% de um intervalo bicaudal exige o valor crítico t de 2,131. O intervalo resultante é

$$\bar{x} \pm t_{0,025,15} \cdot \frac{s}{\sqrt{n}} = 14.532,5 \pm (2,131)\frac{2055,67}{\sqrt{16}}$$

$$= 14.532,5 \pm 1095,2 = (13.437,3,\ 15.627,7)$$

O intervalo é bastante amplo devido tanto ao tamanho pequeno da amostra, como ao valor grande da variabilidade. O limite de confiança inferior de 95% é obtido usando-se − e 1,753 em vez de ± e 2,131, respectivamente.

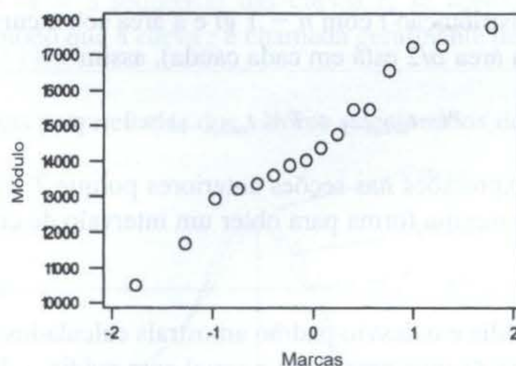


Figura 7.8 Gráfico de probabilidade normal dos dados do módulo de elasticidade ■

Infelizmente, não é fácil selecionar n para controlar a amplitude do intervalo t, porque a amplitude envolve o s desconhecido (antes de os dados serem coletados) e porque n entra não somente por meio de $1/\sqrt{n}$ como também por meio de $t_{\alpha/2,n-1}$. Como resultado, n apropriado pode ser obtido somente por tentativa e erro.

No Capítulo 15, discutiremos um IC de amostra pequena para μ que seja válido somente se a distribuição da população for simétrica, uma suposição mais fraca que a normalidade. Entretanto, quando a distribuição da população é normal, o intervalo t tende a ser menor do que *qualquer* outro intervalo com o mesmo nível de confiança.

Um Intervalo de Previsão para um Único Valor Futuro

Em muitas aplicações, o investigador deseja *prever* um único valor de uma variável a ser observado futuramente, em vez de *estimar* o valor médio dessa variável.

Exemplo 7.12

Considere a seguinte amostra do teor de gordura (em porcentagem) de $n = 10$ cachorros-quentes selecionados aleatoriamente ("Sensory and Mechanical Assessment of the Quality of Frankfurters", *J. Texture Studies,* 1990, p. 395-409):

25,2	21,3	22,8	17,0	29,8	21,0	25,5	16,0	20,9	19,5

Assumindo que foram selecionados, de uma distribuição normal da população, um IC de 95% (estimativa do intervalo de) do teor de gordura médio da população é

$$\bar{x} \pm t_{0,025,9} \cdot \frac{s}{\sqrt{n}} = 21{,}90 \pm 2{,}262 \cdot \frac{4{,}134}{\sqrt{10}} = 21{,}90 \pm 2{,}96$$
$$= (18{,}94,\ 24{,}86)$$

Suponha, entretanto, que você comerá um único cachorro-quente desse tipo e queira uma *previsão* do teor de gordura resultante. A previsão *pontual*, análoga à estimativa *pontual*, é apenas $\bar{x} = 21{,}90$. Essa previsão infelizmente não fornece informações sobre confiabilidade ou precisão. ■

A configuração geral é a seguinte: teremos disponível a amostra aleatória $X_1, X_2, ..., X_n$ de uma distribuição normal da população, e queremos prever o valor de X_{n+1} de uma única observação futura. Um previsor pontual é $\bar{X}$, e o erro de previsão resultante é $\bar{X} - X_{n+1}$. O valor esperado do erro de previsão é

$$E(\bar{X} - X_{n+1}) = E(\bar{X}) - E(X_{n+1}) = \mu - \mu = 0$$

Uma vez que X_{n+1} é independente de $X_1, ..., X_n$, *ele* é independente de $\bar{X}$, de modo que a variância do erro de previsão é

$$V(\bar{X} - X_{n+1}) = V(\bar{X}) + V(X_{n+1}) = \frac{\sigma^2}{n} + \sigma^2 = \sigma^2\left(1 + \frac{1}{n}\right)$$

O erro de previsão é uma combinação linear de vas independentes normalmente distribuídas, de modo que ele próprio seja normalmente distribuído. Dessa forma,

$$Z = \frac{(\bar{X} - X_{n+1}) - 0}{\sqrt{\sigma^2\left(1 + \frac{1}{n}\right)}} = \frac{\bar{X} - X_{n+1}}{\sqrt{\sigma^2\left(1 + \frac{1}{n}\right)}}$$

possui uma distribuição normal padronizada. É possível mostrar que substituindo σ pelo desvio-padrão da amostra S (de $X_1, ..., X_n$), resulta em

$$T = \frac{\bar{X} - X_{n+1}}{S\sqrt{1 + \frac{1}{n}}} \sim \text{distribuição } t \text{ com } n - 1 \text{ gl}$$

Manipular a variável T, como $T = (\bar{X} - \mu)/(S/\sqrt{n})$ foi manipulado no desenvolvimento de um IC, fornece o seguinte resultado.

PROPOSIÇÃO

O intervalo de previsão (IP) de uma única observação a ser selecionada de uma distribuição normal da população é

$$\bar{x} \pm t_{\alpha/2,n-1} \cdot s\sqrt{1 + \frac{1}{n}} \qquad (7.16)$$

O *nível de previsão* é $100(1 - \alpha)\%$.

A interpretação de um nível de previsão de 95% é semelhante à de um nível de confiança de 95%; se o intervalo (7.16) for calculado para amostra após amostra, os 95% de longo prazo desses intervalos incluirão os valores futuros correspondentes de X.

Exemplo 7.13 (continuação do Exemplo 7.12)

Com $n = 10$, $\bar{x} = 21{,}90$, $s = 4{,}134$ e $t_{0{,}025,\,9} = 2{,}262$, um IP de 95% do teor de gordura de um único cachorro-quente é

$$21{,}90 \pm (2{,}262)(4{,}134)\sqrt{1 + \frac{1}{10}} = 21{,}90 \pm 9{,}81$$

$$= (12{,}09,\ 31{,}71)$$

Esse intervalo é bastante amplo, indicando incerteza substancial sobre o teor de gordura. Observe que a amplitude do IP é mais que três vezes a do IC. ■

O erro de previsão é $\bar{X} - X_{n+1}$, diferença entre duas variáveis aleatórias, enquanto o erro de estimativa é $\bar{X} - \mu$, diferença entre uma variável aleatória e um valor fixo (mas desconhecido). O IP é mais amplo que o IC, pois há mais variabilidade no erro de previsão (devido a X_{n+1}) que no erro de estimativa. De fato, à medida que n fica arbitrariamente maior, o IC diminui para o único valor μ e o IP se aproxima de $\mu \pm z_{\alpha/2} \cdot \sigma$. Há incerteza sobre um único valor de X, mesmo quando não há necessidade de fazer estimativa.

Intervalos de Tolerância

Considere uma população de determinado tipo de automóveis e suponha que, sob condições especificadas, o consumo de combustível (milhas/galão) possui distribuição normal com $\mu = 30$ e $\sigma = 2$. Então, uma vez que o intervalo de $-1{,}645$ a $1{,}645$ contendo 90% da área sob a curva z, 90% de todos esses automóveis terão valores de consumo de combustível entre $\mu - 1{,}645\sigma = 26{,}71$ e $\mu + 1{,}645\sigma = 33{,}29$. Mas o que acontece se os valores de μ e σ não são conhecidos? Podemos tomar uma amostra de tamanho n, determinar os consumos de combustível, $\bar{x}$ e s, e formar o intervalo cujo limite inferior seja $\bar{x} - 1{,}645s$ e o limite superior seja $\bar{x} + 1{,}645s$. Entretanto, devido à variabilidade da amostragem nas estimativas de μ e σ, existe uma boa chance de o intervalo resultante incluir menos de 90% dos valores da população. Intuitivamente, para ter uma chance *a priori* de 95%, de o intervalo resultante incluir pelo menos 90% dos valores da população, quando $\bar{x}$ e s são usados no lugar de μ e σ, devemos também substituir 1,645 por algum número maior. Por exemplo: quando $n = 20$, o valor 2,310 é tal que podemos ter 95% de confiança de que o intervalo $\bar{x} \pm 2{,}310s$ incluirá pelo menos 90% dos valores de consumo de combustível na população.

Seja k um número entre 0 e 100. Um **intervalo de tolerância** para incluir pelo menos k% dos valores de uma distribuição normal da população com nível de confiança de 95% possui a forma

$$\bar{x} \pm (\text{valor crítico de tolerância}) \cdot s$$

Os valores críticos de tolerância para $k = 90$, 95 e 99 em combinação com vários tamanhos de amostra são dados na Tabela A.6 do Apêndice. Essa tabela também inclui valores críticos para um nível de confiança de 99% (esses valores são maiores que os valores de 95% correspondentes). Substituir $\pm$ por $+$ fornece um limite de tolerância superior, e usar $-$ no lugar de $\pm$ resulta em um limite de tolerância inferior. Os valores críticos para obter esses limites monocaudais também aparecem na Tabela A.6 do Apêndice.

Exemplo 7.14

Vamos voltar aos dados do módulo de elasticidade discutidos no Exemplo 7.11, onde $n = 16$, $\bar{x} = 14.532{,}5$, $s =$ 2055,67 e um gráfico de probabilidade normal dos dados indicou que a normalidade da população era bastante plausível. Para um nível de confiança de 95%, o intervalo de tolerância bicaudal para incluir pelo menos 95%

dos valores do módulo de elasticidade de espécimes de madeira serrada na população que serviu como amostra utiliza o valor crítico de tolerância de 2,903. O intervalo resultante é

$$14.532,5 \pm (2,903)(2055,67) = 14.532,5 \pm 5967,6 = (8.564,9,\ 20.500,1)$$

Podemos estar altamente confiantes de que pelo menos 95% de todos os espécimes de madeira serrada têm valores de módulo de elasticidade entre 8.564,9 e 20.500,1.

O IC de 95% de μ era (13.437,3, 15.627,7) e o intervalo de previsão de 95% do módulo de elasticidade de um único espécime de madeira serrada é (10.017,0, 19.048,0). O intervalo de previsão e o intervalo de tolerância são substancialmente mais amplos que o intervalo de confiança. ■

Intervalos Baseados em Distribuições Não-normais de População

O IC de t de uma amostra para μ é robusto, para desvios pequenos ou mesmo moderados da hipótese de normalidade, a menos que n seja muito pequeno. Por isso, queremos dizer que, se um valor crítico de confiança de 95%, por exemplo, for usado para calcular o intervalo, o nível de confiança real estará razoavelmente próximo do nível de 95% nominal. Entretanto, se n for pequeno e a distribuição da população for altamente não-normal, então o nível de confiança real pode ser consideravelmente diferente do que você pensa que está utilizando quando obtém um valor crítico específico da tabela de t. Certamente, seria frustrante acreditar que seu nível de confiança é cerca de 95% quando, na verdade, era realmente mais provável ser de 88%! A técnica de *bootstrap*, introduzida na Seção 7.1, tem muito sucesso na estimativa de parâmetros em uma grande variedade de situações não-normais.

Em comparação com o intervalo de confiança, a validade dos intervalos de tolerância e previsão descritos nesta seção está ligada mais proximamente à suposição de normalidade. Esses últimos intervalos não devem ser utilizados na ausência de evidência constrangedora à normalidade. A excelente referência *Statistical Intervals*, citada na bibliografia no final deste capítulo, discute procedimentos alternativos desse tipo para várias outras situações.

Exercícios | Seção 7.3 (28–41)

28. Determine os valores das seguintes quantidades:

a. $t_{0,1,15}$ **b.** $t_{0,05,15}$ **c.** $t_{0,05,25}$ **d.** $t_{0,05,40}$ **e.** $t_{0,005,40}$

29. Determine o valor crítico t que incluirá a área da curva t desejada em cada um dos casos a seguir:

a. Área central = 0,95, gl = 10
b. Área central = 0,95, gl = 20
c. Área central = 0,99, gl = 20
d. Área central = 0,99, gl = 50
e. Área da cauda superior = 0,01, gl = 25
f. Área da cauda inferior = 0,025, gl = 5

30. Determine o valor crítico t de um intervalo de confiança bicaudal em cada uma das situações a seguir:

a. Nível de confiança = 95%, gl = 10
b. Nível de confiança = 95%, gl = 15
c. Nível de confiança = 99%, gl = 15
d. Nível de confiança = 99%, n = 5
e. Nível de confiança = 98%, gl = 24
f. Nível de confiança = 99%, n = 38

31. Determine o valor crítico t de um limite de confiança inferior ou superior para cada uma das situações descritas no Exercício 30.

32. Uma amostra aleatória de n = 8 espécimes de teste de certo tipo de fibra de vidro E produziu uma força do rendimento de corte interfacial média amostral de 30,2 e desvio padrão da amostra de 3,1 ("On Interfacial Failure in Notched Unidirectional Glass/Epoxy Composites", *J. of Composite Materials*, 1985, p. 276-286). Assuma que a força do rendimento de corte interfacial é normalmente distribuída, calcule o IC de 95% para a força média real (como fizeram os autores do artigo citado).

33. O artigo "Measuring and Understanding the Aging of Kraft Insulating Paper in Power Transformers" (*IEEE Electrical Insul. Mag.*, 1996, p. 28-34) continha as seguintes observações sobre o grau de polimerização dos espécimes de papel para os quais a viscosidade vezes a concentração caíram em certo intervalo intermediário:

418	421	421	422	425	427	431
434	437	439	446	447	448	453
454	463	465				

a. Construa um *boxplot* dos dados e comente quaisquer características interessantes.

b. É plausível que as observações da amostra fornecidas tenham sido selecionadas de uma distribuição normal?

c. Calcule o intervalo de confiança de 95% bicaudal para o grau médio real de polimerização (como fizeram os autores do artigo). O intervalo sugere que 440 seja um valor plausível para o grau médio real de polimerização? E quanto a 450?

34. Uma amostra de 14 espécimes de juntas de um tipo específico forneceu o limite de carga proporcional médio da amostra de 8,48 MPa e desvio padrão da amostra de 0,79 MPa ("Characterization of Bearing Strength Factors in Pegged Timber Connections", *J. of Structural Engr.*, 1997, p. 326-332).

a. Calcule e interprete o limite de confiança inferior de 95% para o limite de carga proporcional médio real de todas as juntas. Que hipótese, se houver, você faria sobre a distribuição do limite de carga proporcional?

b. Calcule e interprete o limite de previsão inferior de 95% para o limite de carga proporcional de uma única junta desse tipo.

35. O Exercício 46 no Capítulo 1 introduziu as seguintes observações da amostra sobre viscosidade estabilizada de espécimes de asfalto: 2781, 2900, 3013, 2856, 2888. Um gráfico de probabilidade normal suporta a hipótese de a viscosidade ser distribuída, pelo menos aproximadamente, de forma normal.

a. Calcule a viscosidade média real, de forma que transmita informações sobre precisão e confiabilidade.

b. Preveja a viscosidade de um único espécime de asfalto, de modo que transmita informações sobre precisão e confiabilidade. Como a previsão se compara com a estimativa calculada na parte (a)?

36. As $n = 26$ observações sobre o tempo de fuga dado no Exercício 36 do Capítulo 1 fornecem a média e o desvio padrão da amostra de 370,69 e 24,36, respectivamente.

a. Calcule o limite de confiança superior para o tempo médio de fuga da população, usando um nível de confiança de 95%.

b. Calcule o limite de previsão superior para o tempo de fuga de um único trabalhador adicional, usando um nível de previsão de 95%. Como esse limite se compara ao limite de confiança da parte (a)?

c. Suponha que dois trabalhadores adicionais sejam escolhidos para participar do exercício simulado de fuga. Represente seus tempos de fuga por X_{27} e X_{28}, e por $\overline{X}_{novo}$ a média desses dois valores. Modifique a fórmula de um IP para um único valor de x, a fim de obter o IP de $\overline{X}_{novo}$ e calcule o intervalo bicaudal de 95%, com base nos dados de fuga fornecidos.

37. Um estudo da habilidade de as pessoas andarem em linha reta ("Can We Really Walk Straight?" *Amer. J. of Physical Anthro.*, 1992, p. 19-27) relatou os dados que seguem sobre a cadência (passos por segundo) de uma amostra de $n = 20$ homens saudáveis, aleatoriamente selecionados.

0,95 0,85 0,92 0,95 0,93 0,86 1,00 0,92 0,85 0,81

0,78 0,93 0,93 1,05 0,93 1,06 1,06 0,96 0,81 0,96

Um gráfico de probabilidade normal fornece suporte substancial à hipótese de que a distribuição da população da cadência seja aproximadamente normal. Segue um resumo descritivo dos dados do MINITAB:

```
Variável N  Média  Mediana Média Ap Desv Padrão Média SE
cadência 20 0,9255 0,9300  0,9261      0,0809    0,0181

Variável     Min      Max      Q1      Q3
cadência   0,7800   1,0600   0,8525  0,9600
```

a. Calcule e interprete o intervalo de confiança de 95% para a cadência média da população.

b. Calcule e interprete o intervalo de previsão de 95% para a cadência de uma única pessoa selecionada aleatoriamente dessa população.

c. Calcule o intervalo que inclua pelo menos 99% das cadências da distribuição da população, usando um nível de confiança de 95%.

38. Uma amostra de 25 peças de laminado usado na fabricação de placas de circuito foi selecionada, e a quantidade de deformação (in.) sob condições específicas foi determinada para cada peça, resultando em uma deformação média amostral de 0,0635 e um desvio padrão de 0,0065.

a. Calcule uma previsão para a quantidade de deformação de uma única peça de laminado, de modo que forneça informações sobre precisão e confiabilidade.

b. Calcule o intervalo no qual você possa ter alto grau de confiança de que pelo menos 95% de todas as peças de laminado resultem em quantidades de deformação que estejam entre os dois limites do intervalo.

39. O Exercício 72 do Capítulo 1 forneceu as seguintes observações sobre a medida de absorção do receptor (volume de distribuição ajustado) de uma amostra de 13 pessoas saudáveis: 23, 39, 40, 41, 43, 47, 51, 58, 63, 66, 67, 69, 72.

a. É plausível que a distribuição da população da qual essa amostra foi selecionada seja normal?

b. Calcule o intervalo para o qual possamos ter 95% de confiança de que pelo menos 95% de todas as pessoas saudáveis da população tenham ajustado os volumes da distribuição ajustados incluídos entre os limites do intervalo.

c. Preveja o volume da distribuição ajustado de uma única pessoa saudável, calculando o intervalo de previsão de 95%. Como a amplitude desse intervalo se compara com o intervalo calculado na parte (b)?

40. O Exercício 13 do Capítulo 1 apresentou uma amostra de $n = 153$ observações sobre a resistência à tensão máxima, e o Exercício 17 da seção anterior forneceu quantidades resumidas e pediu o intervalo de confiança de amostra grande. Em virtude de o tamanho da amostra ser

grande, nenhuma suposição sobre a distribuição da população é necessária para a validade do IC.

a. Qualquer suposição sobre a distribuição da resistência à tensão é necessária, antes de calcular um limite de previsão inferior para a resistência à tensão do próximo espécime selecionado usando o método descrito nesta seção? Explique.

b. Use um programa estatístico para investigar a plausibilidade de uma distribuição normal da população.

c. Calcule o limite de previsão inferior com um nível de previsão de 95% para a resistência extrema à tensão do próximo espécime selecionado.

41. Uma tabulação mais extensiva dos valores críticos t que a mostrada neste livro indica que, para a distribuição t com gl 20, as áreas à direita dos valores 0,687, 0,860 e 1,064 são 0,25, 0,20 e 0,15, respectivamente. Qual é o nível de confiança de cada um dos três intervalos de confiança a seguir para a média μ de uma distribuição normal da população? Qual dos três intervalos você recomendaria para ser usado e por quê?

a. $(\bar{x} - 0{,}687s/\sqrt{21}, \bar{x} + 1{,}725s/\sqrt{21})$

b. $(\bar{x} - 0{,}860s/\sqrt{21}, \bar{x} + 1{,}325s/\sqrt{21})$

c. $(\bar{x} - 1{,}064s/\sqrt{21}, \bar{x} + 1{,}064s/\sqrt{21})$

7.4 Intervalos de Confiança para Variância e Desvio Padrão de uma População Normal

Embora as inferências relacionadas à variância de uma população σ^2 ou ao desvio padrão σ sejam geralmente de menor interesse que as referentes à média ou proporção, há ocasiões em que tais procedimentos são necessários. No caso de uma distribuição normal da população, as inferências são feitas com base no seguinte resultado relacionado à variância da amostra S^2.

TEOREMA

Seja $X_1, X_2, ..., X_n$ a amostra aleatória de uma distribuição normal com parâmetros μ e σ^2. Então, a va

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X_i - \bar{X})^2}{\sigma^2}$$

possui distribuição de probabilidade qui-quadrada (χ^2) com $n - 1$ gl.

Conforme discutido nas seções 4.4 e 7.1, a distribuição qui-quadrada é uma distribuição de probabilidade contínua com um único parâmetro ν, chamado número de graus de liberdade, com valores possíveis 1, 2, 3, ... Os gráficos de diversas funções-distribuição de probabilidade de χ^2 (fdps) são ilustrados na Figura 7.9. Cada fdp $f(x; \nu)$ é positiva somente para $x > 0$, e cada uma possui inclinação positiva (cauda superior longa), embora a distribuição se mova à direita e se torne mais simétrica à medida que ν aumenta. Para especificar os procedimentos inferenciais que utilizam a distribuição qui-quadrada, precisamos de notação análoga à de um valor crítico t de $t_{\alpha,\nu}$.

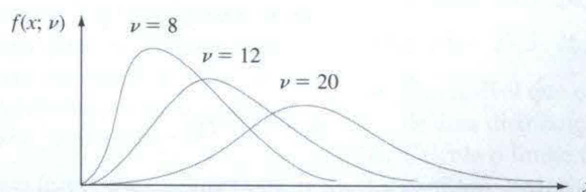


Figura 7.9 Gráficos das funções-densidade qui-quadrada

> **Notação**
>
> Seja $\chi^2_{\alpha,\nu}$, chamado **valor crítico qui-quadrado,** o número no eixo da medição, de modo que α da área sob a curva qui-quadrada com gl ν esteja à direita de $\chi^2_{\alpha,\nu}$.

Em virtude de a distribuição t ser simétrica, foi necessário tabular somente os valores críticos da cauda superior ($t_{\alpha,\nu}$ para valores pequenos de α). A distribuição qui-quadrada não é simétrica; de modo que, a Tabela A.7 do Apêndice contém valores de $\chi^2_{\alpha,\nu}$, tanto para α próximo de 0 e de 1, conforme ilustrado na Figura 7.10(b). Por exemplo: $\chi^2_{0,025,14} = 26,119$ 4 $\chi^2_{0,95,20}$ (o 5º percentil) $= 10,851$.

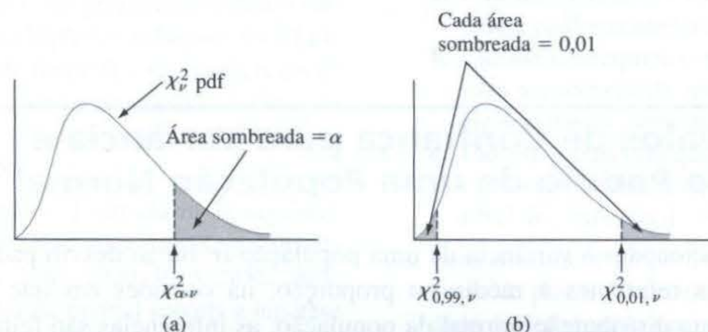


Figura 7.10 Notação ilustrada de $\chi^2_{\alpha,\nu}$

A va $(n-1)S^2/\sigma^2$ satisfaz as duas propriedades nas quais o método geral para obter IC se baseia: é uma função do parâmetro de interesse σ^2, ainda que sua distribuição de probabilidade (qui-quadrada) não dependa desse parâmetro. A área sob uma curva qui-quadrada com gl ν à direita de $\chi^2_{\alpha/2,\nu}$ é $\alpha/2$, como a área à esquerda de $\chi^2_{1-\alpha/2,\nu}$. Dessa forma, a área incluída entre esses dois valores críticos é $1-\alpha$. Como conseqüência disso e do teorema que acabamos de enunciar,

$$P\left(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}\right) = 1-\alpha \qquad (7.17)$$

As desigualdades de (7.17) são equivalentes a

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

Substituindo o valor calculado s^2 nos limites fornece o IC para σ^2 e extraindo as raízes quadradas obtém-se o intervalo de σ.

> **O intervalo de confiança de 100(1 − α)% para a variância σ^2 de uma população normal** possui limite inferior
>
> $$(n-1)s^2/\chi^2_{\alpha/2,n-1}$$
>
> e limite superior
>
> $$(n-1)s^2/\chi^2_{1-\alpha/2,n-1}$$
>
> **O intervalo de confiança para σ** possui limites superior e inferior que são as raízes quadradas dos limites correspondentes no intervalo para σ^2.

Exemplo 7.15

Os dados que seguem sobre a voltagem de quebra de circuitos carregados eletricamente foram obtidos de um gráfico de probabilidade normal que apareceu no artigo "Damage of Flexible Printed Wiring Boards Associated with Lightning-Induced Voltage Surges" (*IEEE Transactions on Components, Hybrids, and Manuf. Tech.*, 1985, p. 214-220). A disposição em reta do gráfico forneceu forte apoio à suposição de que a voltagem de quebra é aproximadamente distribuída de forma normal.

1470 1510 1690 1740 1900 2000 2030 2100 2190

2200 2290 2380 2390 2480 2500 2580 2700

Seja σ^2 a variância da distribuição da voltagem de quebra. O valor calculado da variância da amostra é $s^2 = 137.324,3$, a estimativa pontual de σ^2. Com gl $= n - 1 = 16$, o IC de 95% requer que $\chi^2_{0,975,16} = 6,908$ e $\chi^2_{0,025,16} = 28,845$. O intervalo é

$$\left(\frac{16(137.324,3)}{28,845}, \frac{16(137.324,3)}{6,908}\right) = (76.172,3, 318.064,4)$$

Extraindo a raiz quadrada de cada ponto final produz (276,0, 564,0), IC de 95% para σ. Esses intervalos são bastante amplos, refletindo a variabilidade substancial na voltagem de quebra em combinação com o tamanho pequeno de uma amostra. ■

Os ICs de σ^2 e σ, quando a distribuição da população não é normal, podem ser difíceis de se obter, mesmo quando o tamanho da amostra é grande. Para esses casos, consulte um estatístico de renome.

Exercícios | Seção 7.4 (42–46)

42. Determine os valores das seguintes quantidades:
   a. $\chi^2_{0,1.15}$ b. $\chi^2_{0,1.25}$
   c. $\chi^2_{0,01.25}$ d. $\chi^2_{0,005.25}$
   e. $\chi^2_{0,99.25}$ f. $\chi^2_{0,995.25}$

43. Determine o seguinte:
   a. O 95º percentil da distribuição qui-quadrada com $\nu = 10$
   b. O 5º percentil da distribuição qui-quadrada com $\nu = 10$
   c. $P(10,98 \leq \chi^2 \leq 36,78)$, onde χ^2 é uma va qui-quadrada com $\nu = 22$
   d. $P(\chi^2 < 14,611$ ou $\chi^2 > 37,652)$, onde χ^2 é uma va qui-quadrada com $\nu = 25$

44. A quantidade de expansão lateral (mils) foi determinada para uma amostra de $n = 9$ soldas de arco de gás metálico acionado por pulsos usados em tanques de carga de navio LNG. O desvio padrão da amostra resultante foi $s = 2,81$ mils. Assumindo a normalidade, encontre o IC de 95% para σ^2 e para σ.

45. As seguintes observações foram feitas sobre a dureza de fratura de uma placa de base de 18% de níquel em aço inoxidável ["Fracture Testing of Weldments", *ASTM Special Publ.* nº 381, 1965, p. 328-356 (in ksi $\sqrt{\text{in.}}$, dado em ordem crescente)]:

69,5 71,9 72,6 73,1 73,3 73,5 75,5 75,7

75,8 76,1 76,2 76,2 77,0 77,9 78,1 79,6

79,7 79,9 80,1 82,2 83,7 93,7

Calcule o IC de 99% para o desvio padrão da distribuição de dureza de fratura. Esse intervalo é válido qualquer que seja a natureza da distribuição? Explique.

46. Os resultados de um teste de turvação de Wagner realizado em 15 amostras de areia de teste-padrão de Otawa foram (em microampères):

26,7 25,8 24,0 24,9 26,4 25,9 24,4 21,7

24,1 25,9 27,3 26,9 27,3 24,8 23,6

   a. É plausível que essa amostra tenha sido selecionada de uma distribuição normal da população?
   b. Calcule o limite de confiança superior com nível de confiança de 95% para o desvio padrão da população de turvação.

Exercícios Suplementares (47–62)

47. O Exemplo 1.11 introduziu as observações que seguem sobre resistência de contato.

11,5	12,1	9,9	9,3	7,8	6,2	6,6	7,0
13,4	17,1	9,3	5,6	5,7	5,4	5,2	5,1
4,9	10,7	15,2	8,5	4,2	4,0	3,9	3,8
3,6	3,4	20,6	25,5	13,8	12,6	13,1	8,9
8,2	10,7	14,2	7,6	5,2	5,5	5,1	5,0
5,2	4,8	4,1	3,8	3,7	3,6	3,6	3,6

a. Estime a resistência de contato média real, de modo que transmita informações sobre precisão e confiabilidade. (*Sugestão*: $\sum x_i = 387{,}8$ e $\sum x_i^2 = 4247{,}08$.)
b. Calcule o IC de 95% para a proporção de todas as junções cujos valores de resistência excederiam 10.

48. Um triatlo que consiste em nadar, andar de bicicleta e correr é um dos eventos de esporte amador mais árduos. O artigo "Cardiovascular and Thermal Response of Triathlon Performance" (*Medicine and Science in Sports and Exercise,* 1988, p. 385-389) relata um estudo de pesquisa que envolve nove triatletas do sexo masculino. O índice cardíaco máximo (batimentos/min) foi registrado durante a realização de cada um dos três eventos. Para a natação, a média e o desvio padrão amostrais foram 188,0 e 7,2, respectivamente. Assumindo que a distribuição do índice cardíaco seja (aproximadamente) normal, construa um IC de 98% para o índice cardíaco médio real dos triatletas durante a natação.

49. Para cada um dos 18 centros preservados de reservatórios de carbonato de óleo, a quantidade de saturação de gás residual após a injeção de um solvente foi medida na saída do fluxo de água. As observações, em porcentagem de volume de poros, foram

23,5	31,5	34,0	46,7	45,6	32,5
41,4	37,2	42,5	46,9	51,5	36,4
44,5	35,7	33,5	39,3	22,0	51,2

(Consulte "Relative Permeability Studies of Gas-Water Flow Following Solvent Injection in Carbonate Rocks," *Soc. Petroleum Engineers J.*, 1976, p. 23-30.)
a. Construa um *boxplot* desses dados e faça comentários sobre quaisquer características interessantes.
b. É plausível que a amostra fosse selecionada de uma distribuição normal da população?
c. Calcule o IC de 98% para a quantidade média real de saturação de gás residual.

50. Um artigo de um periódico relata que uma amostra de tamanho 5 foi usada como base para calcular o IC de 95% para a freqüência natural média real (Hz) de vigas laminadas de determinado tipo. O intervalo resultante foi (229,764, 233,504). Você decide que um nível de confiança de 99% é mais apropriado que o nível de 95% utilizado. Quais são os limites do intervalo de 99%? (*Sugestão*: use o centro do intervalo e sua amplitude para determinar $\bar{x}$ e s.)

51. O gerente financeiro de uma grande rede de lojas de departamentos selecionou uma amostra aleatória de 200 de seus clientes com cartão de crédito e descobriu que 136 pagaram juros no ano anterior devido a saldo não-pago.
a. Calcule o IC de 90% para a proporção real dos clientes com cartão de crédito que pagaram juros no ano anterior.
b. Se a amplitude desejada do intervalo de 90% for 0,05, que tamanho de amostra é necessário para garantir isso?
c. O limite superior do intervalo na parte (a) especifica um limite de confiança superior de 90% para a proporção que está sendo estimada? Explique.

52. O tempo de reação a um estímulo compreende o intervalo de tempo entre o início e o fim desse estímulo, sinalizado pelo primeiro movimento visível. O artigo "Relationship of Reaction Time and Movement Time in a Gross Motor Skill" (*Perceptual and Motor Skills,* 1973, p. 453-454) relata que o TR médio da amostra de 16 nadadores experientes a um tiro inicial foi de 0,214 s. e o desvio padrão da amostra foi de 0,036 s.
a. Fazendo quaisquer suposições necessárias, encontre o IC de 90% do TR médio real de todos os nadadores experientes.
b. Calcule o limite de confiança superior de 90% para o desvio padrão da distribuição do tempo de reação.
c. Preveja o TR para outro indivíduo, de forma que transmita informações sobre precisão e confiabilidade.

53. A infestação de pulgão em árvores frutíferas pode ser controlada pela pulverização de pesticida ou inundação de joaninhas. Em uma área específica, quatro pomares de diferentes árvores frutíferas são selecionados para experiência. Os três primeiros pomares são pulverizados com pesticidas 1, 2 e 3, respectivamente, e o quarto é tratado com joaninhas, produzindo os seguintes resultados:

Tratamento	n_i = número de árvores	$\bar{x}_i$ (alqueires/árvore)	s_i
1	100	10,5	1,5
2	90	10,0	1,3
3	100	10,1	1,8
4	120	10,7	1,6

Seja μ_i = o rendimento médio real (alqueires/árvore) depois de receber o i-ésimo tratamento. Então

$$\theta = \frac{1}{3}(\mu_1 + \mu_2 + \mu_3) - \mu_4$$

meça a diferença nos rendimentos médios reais entre o tratamento com pesticidas e o tratamento com joaninhas. Quando n_1, n_2, n_3 e n_4 são grandes, o estimador $\hat{\theta}$ obtido pela substituição de cada μ_i por $\overline{X}_i$ é aproximadamente normal. Use esses dados para determinar o IC de $100(1 - \alpha)\%$ de amostra grande para θ e calcule o intervalo de 95% para os dados fornecidos.

54. É importante que as máscaras usadas pelos bombeiros sejam capazes de resistir a altas temperaturas, pois esses profissionais trabalham com freqüência em temperaturas de 200-500°F. Em um teste de um tipo de máscara, 11 dos 55 equipamentos tiveram as lentes estouradas a 250°. Construa o IC de 90% para a proporção real de máscaras desse tipo, cujas lentes estourariam a 250°.

55. Um fabricante de livros didáticos está interessado em estimar a resistência das capas produzidas por uma máquina de encadernação específica. A resistência pode ser medida registrando-se a força necessária para arrancar as páginas da encadernação. Se essa força for medida em libras, quantos livros devem ser testados para estimar a força média necessária para quebrar a encadernação dentro de 0,1 lb com confiança de 95%? Assuma que σ seja 0,8.

56. A exposição crônica à fibra de amianto é um perigo à saúde bastante conhecido. O artigo "The Acute Effects of Chrysotile Asbestos Exposure on Lung Function" (*Environ. Research*, 1978, p. 360-372) relata resultados de um estudo com base em uma amostra de operários de construção expostos ao amianto durante um longo período. Dentre os dados fornecidos no artigo, 8 meses após o período de exposição, os seguintes valores (ordenados) de complacência pulmonar ($cm^3/cm\ H_2O$) foram apresentados em cada uma das 16 pessoas (a complacência pulmonar é uma medida da elasticidade do pulmão, ou como os pulmões, de maneira eficaz, são capazes de inalar e exalar):

167,9	180,8	184,8	189,8	194,8	200,2
201,9	206,9	207,2	208,4	226,3	227,7
228,5	232,4	239,8	258,6		

a. É plausível que a distribuição da população seja normal?

b. Calcule o IC de 95% para a complacência pulmonar média real após a exposição.

c. Calcule o intervalo que, com um nível de confiança de 95%, inclua pelo menos 95% dos valores de complacência pulmonar na distribuição da população.

57. No Exemplo 6.7, introduzimos o conceito de um experimento censurado em que n componentes são testados e que o experimento termina logo que r dos componentes tiver falhado. Suponha que a vida útil dos componentes seja independente, cada um com uma distribuição exponencial com parâmetro λ. Represente por Y_1 a hora em que ocorre a primeira falha, por Y_2 a hora em que ocorre a segunda falha, e assim por diante, de modo que $T_r = Y_1 + \cdots + Y_r + (n - r)Y_r$ seja a vida útil total acumulada no fim. Então, é possível mostrar que $2\lambda T_r$ possui uma distribuição qui-quadrada com gl de $2r$. Use esse fato para desenvolver uma fórmula de IC de $100(1 - \alpha)\%$ da vida útil média real de $1/\lambda$. Calcule o IC de 95% a partir dos dados no Exemplo 6.7.

58. Seja por $X_1, X_2, ..., X_n$ a amostra aleatória de uma distribuição de probabilidade contínua que tem mediana $\tilde{\mu}$ (de modo que $P(X_i \le \tilde{\mu}) = P(X_i \ge \tilde{\mu}) = 0{,}5$).

a. Mostre que

$$P(\text{mín}(X_i) < \tilde{\mu} < \text{máx}(X_i)) = 1 - \left(\frac{1}{2}\right)^{n-1}$$

para que $(\text{mín}(x_i), \text{máx}(x_i))$ seja um intervalo de confiança de $100(1 - \alpha)\%$ de $\tilde{\mu}$ com $\alpha = \left(\frac{1}{2}\right)^{n-1}$. [*Sugestão*: o complemento do evento $\{\text{mín}(X_i) < \tilde{\mu} < \text{máx}(X_i)\}$ é $\{\text{máx}(X_i) \le \tilde{\mu}\} \cup \{\text{mín}(X_i) \ge \tilde{\mu}\}$. Mas $\text{máx}(X_i) \le \tilde{\mu}$ se é somente se $X_i \le \tilde{\mu}$ para todos os i.]

b. Para cada um dos seis meninos normais, a quantidade do aminoácido alanina (mg/100 mL) foi determinada enquanto as crianças estavam em uma dieta sem isoleucina, resultando nos seguintes dados:

2,84 3,54 2,80 1,44 2,94 2,70

Calcule o IC de 97% da quantidade mediana real de alanina das crianças nessa dieta ("The Essential Amino Acid Requirements of Infants," *Amer. J. Nutrition*, 1964, p. 322-330).

c. Seja $x_{(2)}$ o segundo menor x_i e $x_{(n-1)}$ o segundo maior x_i. Qual é o coeficiente de confiança do intervalo $(x_{(2)}, x_{(n-1)})$ para $\tilde{\mu}$?

59. Seja $X_1, X_2, ..., X_n$ a amostra aleatória de uma distribuição uniforme no intervalo $[0, \theta]$, de modo que

$$f(x) = \begin{cases} \dfrac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{caso contrário} \end{cases}$$

Então, se $Y = \text{máx}(X_i)$, é possível mostrar que a va $U = Y/\theta$ possui função densidade

$$f_U(u) = \begin{cases} nu^{n-1} & 0 \le u \le 1 \\ 0 & \text{caso contrário} \end{cases}$$

a. Use $f_U(u)$ para verificar que

$$P\left((\alpha/2)^{1/n} < \frac{Y}{\theta} \le (1 - \alpha/2)^{1/n}\right) = 1 - \alpha$$

e use a expressão para encontrar o IC de $100(1 - \alpha)\%$ para θ.

b. Verifique se $P(\alpha^{1/n} \le Y/\theta \le 1) = 1 - \alpha$ e encontre o IC de $100(1 - \alpha)\%$ para θ com base nessa declaração de probabilidade.

c. Qual dos dois intervalos deduzidos anteriormente é menor? Se meu tempo de espera por um ônibus, pela manhã, for uniformemente distribuído e os tempos de espera observados forem $x_1 = 4{,}2$, $x_2 = 3{,}5$, $x_3 = 1{,}7$, $x_4 = 1{,}2$, e $x_5 = 2{,}4$, encontre o IC de 95% para θ usando o intervalo menor.

60. Seja $0 \leq \gamma \leq \alpha$. Então, o IC de $100(1-\alpha)$ % para μ quando n for grande é

$$\left(\bar{x} - z_{\gamma} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha-\gamma} \cdot \frac{s}{\sqrt{n}}\right)$$

A escolha $\gamma = \alpha/2$ produz o intervalo usual encontrado na Seção 7.2; se $\gamma \neq \alpha/2$, esse intervalo não é simétrico em torno de $\bar{x}$. A amplitude desse intervalo é $w = s(z_{\gamma} + z_{\alpha-\gamma})/\sqrt{n}$. Mostre que w é minimizado para $\gamma = \alpha/2$, de modo que o intervalo simétrico é o menor. [*Sugestões*: (a) pela definição de z_{α}, $\Phi(z_{\alpha}) = 1 - \alpha$, de modo que $z_{\alpha} = \Phi^{-1}(1-\alpha)$; (b) a relação entre a derivada de uma função $y = f(x)$ e a função inversa $x = f^{-1}(y)$ é $(d/dy) f^{-1}(y) = 1/f'(x)$.]

61. Suponha que $x_1, x_2, \ldots, x_n$ sejam valores observados que resultam de uma amostra aleatória de distribuição simétrica, porém possivelmente de cauda longa. Sejam $\tilde{x}$ e f_s a mediana da amostra e a quarta dispersão, respectivamente. O Capítulo 11 de *Understanding Robust and Exploratory Data Analysis* (consulte a bibliografia do Capítulo 6) sugere o seguinte IC robusto de 95% para a média da população (ponto de simetria):

$$\tilde{x} \pm \left(\frac{\text{valor crítico de } t \text{ conservativo}}{1{,}075}\right) \cdot \frac{f_s}{\sqrt{n}}$$

O valor da quantidade entre parênteses é 2,10 para $n = 10$, 1,94 para $n = 20$ e 1,91 para $n = 30$. Calcule esse IC para os dados do Exercício 45, e compare ao IC de t apropriado para a distribuição de uma população normal.

62. **a.** Use os resultados do Exemplo 7.5 para obter o limite de confiança inferior de 95% do parâmetro λ de uma distribuição exponencial, e calcule o limite com base nos dados fornecidos no exemplo.

b. Se a vida útil X tem uma distribuição exponencial, a probabilidade de a vida útil exceder t é $P(X > t) = e^{-\lambda t}$. Use o resultado da parte (a) para obter o limite de confiança inferior de 95% para a probabilidade de o tempo de quebra exceder 100 min.

Bibliografia

DEGROOT, Morris e SCHERVISH, Mark, *Probability and Statistics*, Reading, MA: Addison-Wesley, 2002. (3. ed.) Uma exposição muito boa dos princípios gerais de inferência estatística.

HAHN, Gerald e MEEKER, William, *Statistical Intervals*, Wiley: Nova York, 1991. Tudo que você sempre quis saber sobre os intervalos estatísticos (confiança, previsão, tolerância e outros).

LARSEN, Richard e MARX, Morris, *Introduction to Mathematical Statistics* (2. ed.), Prentice Hall, Englewood Cliffs, NJ: 1986. Semelhante à apresentação de DeGroot, mas levemente menos matemático.

Respostas dos Exercícios de Numeração Ímpar

Capítulo 1

1. a. *Los Angeles Times, Oberlin Tribune, Gainesville Sun, Washington Post*
b. Duke Energy, Clorox, Seagate, Neiman Marcus
c. Vince Correa, Catherine Miller, Michael Cutler, Ken Lee
d. 2,97; 3,56; 2,20; 2,97

3. a. Qual a probabilidade de, se mais da metade dos VCRs amostrados precisarem ou precisaram do serviço de garantia? Qual é o número esperado entre os 100 que precisarem do serviço de garantia? Qual a probabilidade disso se o número dos que estão necessitando de serviço de garantia exceder o número esperado em mais do que 10?
b. Suponha que 15 de 100 amostrados precisaram do serviço de garantia. Quão confiantes podemos estar de que a proporção de *todos* os VCRs que precisam de serviço de garantia estão entre 0,08 e 0,22? A amostra fornece evidência incontestável que permite concluir que mais do que 10% de todos VCR's precisam de serviço de garantia?

5. a. Não. Todos os estudantes que fizeram um grande curso de estatística foram os que participaram de um programa de SI deste tipo.
b. A aleatoriedade protege contra vários desvios e ajuda a garantir que todos os do grupo SI sejam tão semelhantes quanto os estudantes do grupo de controle.
c. Não haveria bases sólidas para o cálculo efetivo do SI (nenhum dos quais a pontuação SI poderia ser razoável quando comparada).

7. Pode-se gerar uma amostra aleatória simples de todas as casas de família na cidade ou uma amostra aleatória estratificada tomando uma aleatória simples de cada um dos 10 bairros vizinhos. A partir de cada casa selecionada, os valores de todas as variáveis desejadas deveriam ser determinados. Isso seria um estudo enumerativo, pois existe uma população de objetos finita e identificável a partir da qual amostrar.

9. a. Possivelmente erro de medida, erro de registro, diferenças nas condições ambientais no tempo de medida etc.
b. Não. Não há quadro de amostragem.

11.

6L	430
6H	769689
7L	42014202
7H	
8L	011211410342
8H	9595578
9L	30
9H	58

A lacuna nos dados — sem contagem acima de 70.

13. a. 12 2 folha: dígito das unidades
12 445
12 6667777
12 889999
13 00011111111
13 222222222222333333333333
13 44444444444444444455555555555555555555
13 6666666666666677777777
13 888888888888899999
14 0000001111
14 2333333
14 444
14 77
simetria
b. Próximo do formato de sino, centro ≈ 135, dispersão significativa, sem lacunas ou outliers.

15.

Crocante		Cremoso
	2	2
644	3	0069
77220	4	00145
6320	5	003666
222	6	258
55	7	
0	8	

caule: dezenas
folha: unidades

Ambos os conjuntos de contagens são muito dispersos. Parece não haver outliers. A distribuição de contagens de crocantes parece estar deslocada para a direita (na direção dos valores maiores) do que para contagens de cremosas por algo na ordem de 10.

17. a.

# Não-conformidade	Freqüência	Freq. Rel.
0	7	0,117
1	12	0,200
2	13	0,217
3	14	0,233
4	6	0,100
5	3	0,050
6	3	0,050
7	1	0,017
8	1	0,017
	60	1,001

b. 0,917, 0,867, 1 − 0,867 = 0,133

c. O histograma tem uma inclinação substancialmente positiva. Está centrado em algum lugar entre 2 e 3 e dispersa-se um pouco em seu centro.

19. a. 0,99 (99%), 0,71 (71%)

b. 0,64 (64%), 0,44 (44%)

c. Estritamente falando, o histograma não é unimodal, mas está próximo de ser, portanto com uma inclinação positiva moderada. Um tamanho de amostra muito maior resultaria provavelmente em uma figura mais plana.

21. a.

y	Freq.	Freq. Rel.
0	17	0,362
1	22	0,468
2	6	0,128
3	1	0,021
4	0	0,000
5	1	0,021
	47	1,000

0,362; 0,638

b.

z	Freq.	Freq. Rel.
0	13	0,277
1	11	0,234
2	3	0,064
3	7	0,149
4	5	0,106
5	3	0,064
6	3	0,064
7	0	0,000
8	2	0,043
	47	1,001

0,894; 0,830

23. a.

Classe	Freq.	Freq. Rel.
0–<100	21	0,21
100–<200	32	0,32
200–<300	26	0,26
300–<400	12	0,12
400–<500	4	0,04
500–<600	3	0,03
600–<700	1	0,01
700–<800	0	0,00
800–<900	1	0,01
	100	1,00

b.

Classe	Freq.	Freq. Rel.	Densidade
0–<50	8	0,08	0,0016
50–<100	13	0,13	0,0026
100–<150	11	0,11	0,0022
150–<200	21	0,21	0,0042
200–<300	26	0,26	0,0026
300–<400	12	0,12	0,0012
400–<500	4	0,04	0,0004
500–<600	3	0,03	0,0003
600–<900	2	0,02	0,00007
	100	1,00	

c. 0,79

25.

Classe	Freq.	Classe	Freq.
10–<20	8	1,1–<1,2	2
20–<30	14	1,2–<1,3	6
30–<40	8	1,3–<1,4	7
40–<50	4	1,4–<1,5	9
50–<60	3	1,5–<1,6	6
60–<70	2	1,6–<1,7	4
70–<80	1	1,7–<1,8	5
	40	1,8–<1,9	1
			40

Original: inclinada positivamente;
Transformada: muito mais simétrica, não distante da forma de sino.

27. a. A observação 50 cai em uma classe limite.

b.

Classe	Freq.	Freq. Rel.
0–<50	9	0,18
50–<100	19	0,38
100–<150	11	0,22
150–<200	4	0,08
200–<300	4	0,08
300–<400	2	0,04
400–<500	0	0,00
500–<600	1	0,02
	50	1,00

Um valor (central) representativo é tanto um pouco abaixo como um pouco acima de 100, dependendo de como se mede o centro. Há grande variação na vida útil, especialmente nos valores na extremidade superior dos dados. Há vários candidatos à outliers.

c.

Classe	Freq.	Freq. Rel.
2,25–<2,75	2	0,04
2,75–<3,25	2	0,04
3,25–<3,75	3	0,06
3,75–<4,25	8	0,16
4,25–<4,75	18	0,36
4,75–<5,25	10	0,20
5,25–<5,75	4	0,08
5,75–<6,25	3	0,06
	50	1,00

Há muito mais simetria na distribuição dos valores de $\ln(x)$ do que nos valores x em si mesmos, e menos variabilidade. Não há lacunas grandes ou óbvios outliers.

d. 0,38; 0,14.

29.

Reclamação	Freq.	Freq. Rel.
J	10	0,1667
F	9	0,1500
B	7	0,1167
M	4	0,0667
C	3	0,0500
N	6	0,1000
O	21	0,3500
	60	1,0001

31.

Classe	Freq.	Freq. Acum.	Freq. Rel. Acum.
0–<4	2	2	0,050
4–<8	14	16	0,400
8–<12	11	27	0,675
12–<16	8	35	0,875
16–<20	4	39	0,975
20–<24	0	39	0,975
24–<28	1	40	1,000

33. **a.** $\bar{x} = 192{,}57$, $\tilde{x} = 189{,}0$
b. Novo $\bar{x} = 189{,}71$; $\tilde{x}$ inalterado
c. 191,0; 7,14% **d.** 122,6

35. **a.** $\bar{x} = 12{,}55$, $\tilde{x} = 12{,}50$, $\bar{x}_{tr(12,5)} = 12{,}40$. Deleção da maior observação (18,0) faz com que $\tilde{x}$ e $\bar{x}_{tr}$ sejam um pouco menores do que $\bar{x}$.
b. Por no máximo 4.0
c. Não; multiplique os valores de $\bar{x}$ e $\tilde{x}$ pelo fator de conversão 1/2,2.

37. $\bar{x}_{tr(10)} = 11{,}46$

39. **a.** $\bar{x} = 1{,}0297$, $\tilde{x} = 1{,}009$ **b.** 0,383

41. **a.** 0,7 **b.** Também 0,7 **c.** 13

43. $\tilde{x} = 68{,}0$, $\bar{x}_{tr(20)} = 66{,}2$, $\bar{x}_{tr(30)} = 67{,}5$

45. **a.** $\bar{x} = 115{,}58$; os desvios são 0,82; 0,32; −0,98; −0,38; 0,22
b. 0,482; 0,694 **c.** 0,482 **d.** 0,482

47. $\bar{x} = 116{,}2$, $s = 25{,}75$. A magnitude de s indica uma quantidade substancial de variação em torno do centro (um desvio "representativo" de aproximadamente 25).

49. **a.** 0,56,80; 197,8040 **b.** 0,5016; 0,708

51. **a.** 1264,766; 35,564 **b.** 0,351; 0,593

53. **a.** 2,74; 3,88 **b.** 1,14 **c.** Inalterado
d. No máximo 0,40 **e.** 1,19

55. **a.** 33 **b.** Não
c. Inclinação positiva leve na metade do meio, mas simétrica no todo. A extensão da variabilidade parece significativa
d. No máximo 32

57. **a.** Sim. 125,8 é um outlier extremo e 250,2 é uma outlier moderado.
b. Além da presença de outliners, há inclinações positivas ambas em 50% do meio dos dados e, excetuando os outliers, em todo ele. Exceto para os dois outliers, parece haver uma quantidade relativamente pequena de variabilidade nos dados.

59. **a.** ED: 0,4; 0,10; 2,75; 2,65;
Não-Ed: 1,60; 0,30; 7,90; 7,60
b. ED: 8,9 e 9,2 são outliers moderados e 11,7 e 21,0 são outliers extremos.
Não há outliners na amostra não-ED.
c. Quatro outliers para ED, nenhum para não-ED. Inclinação positiva significativa em ambas as amostras; menos variabilidade em ED (o menor f_s) e observações não-ED tendem a ser algo maiores do que as observações ED.

61. Outliers, ambos moderados e extremos, apenas às 6 da manhã. As distribuições em outros horários são bem simétricas. A variabilidade aumenta consideravelmente até 14h e então diminui ligeiramente, e o mesmo acontece para os valores de coeficiente gasolina-vapor "típicos".

63. Baseadas em um boxplot comparativa, as três amostras parecem significativamente diferentes em relação ao centro. Há muito menos variabilidade para a taxa de fluxo de 160 do que para as duas outras taxas. Nenhum outlier presente. Há evidência de inclinação positiva moderada nos 50% do meio de cada amostra, mas não para a amostra da taxa de fluxo de 160 ou a amostra da taxa de fluxo de 200 como um todo.

65. **a.** 9,59; 59,41
b. $CV = 0{,}396$ para HC e $= 0{,}323$ para CO, portanto há maior variabilidade relativa nos dados de HC ainda que seu DP seja muito menor.

67. 10,65

69. **a.** $\bar{y} = a\bar{x} + b$, $s_y^2 = a^2 s_x^2$ **b.** 189,14; 1,87

71. **a.** A média, a mediana e a média ordenada são virtualmente idênticas, sugerindo uma quantidade substancial de simetria nos dados: o fato de os quartis terem aproximadamente as mesmas distâncias da mediana e de as menores e maiores observações serem aproximadamente eqüidistantes do centro fornece suporte adicional para a simetria. O desvio padrão é bem pequeno em relação à média e à mediana.
b. Veja os comentários de (a). Além disso, utilizando 1,5(Q3 − Q1) como padrão de medida, as duas maiores e as três menores observações são outliers moderados.

73. $\bar{x} = 0{,}9255$, $s = 0{,}0809$, $\tilde{x} = 0{,}93$, quantidade pequena de variabilidade, um pouco de inclinação.

75. **a.** Os "resumos de cinco números" ($\tilde{x}$, os dois quartos e a menor e a maior observações) são idênticos e não há outliers, portanto os três *boxplots* individuais são idênticos.
b. Diferenças na variabilidade, natureza das lacunas e existência de agrupamentos para as três amostras.
c. Não. Detalhe foi perdido.

77. a.

0	2355566777888	
1	0000135555	
2	00257	
3	0033	
4	0057	caule: unidades
5	044	folha: dezenas
6		
7	05	
8	8	
9	0	
10	3	
HI	22,0; 24,5	

b.

Classe	Freq.	Freq. Rel.	Densidade
0–<2	23	0,500	0,250
2–<4	9	0,196	0,098
4–<6	7	0,152	0,076
6–<10	4	0,087	0,022
10–<20	1	0,022	0,002
20–<30	2	0,043	0,004

79. a. $\bar{x}_{n+1} = (n\bar{x}_n + x_{n+1})/(n+1)$
c. 12,53; 0,532

81. Uma inclinação substancial positiva (assumindo unimodalidade)

83. a. Todos os pontos caem em uma reta de 45°. Os pontos caem abaixo de uma reta de 45°.
b. Os pontos caem bem abaixo de uma reta de 45° indicando uma inclinação substancial positiva.

Capítulo 2

1. a. $\mathcal{S}$ = {1324, 3124, 1342, 3142, 1423, 1432, 4123, 4132, 2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231}
b. A = {1324, 1342, 1423, 1432}
c. B = {2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231}
d. $A \cup B$ = {1324, 1342, 1423, 1432, 2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231},
$A \cap B$ não contém saídas (A e B são disjuntos),
A' = {3124, 3142, 4123, 4132, 2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231}

3. a. A = {SSF, SFS, FSS}
b. B = {SSF, SFS, FSS, SSS}
c. C = {SFS, SSF, SSS}
d. C' = {FFF, FSF, FFS, FSS, SFF},
$A \cup C$ = {SSF, SFS, FSS, SSS},
$A \cap C$ = {SSF, SFS},
$B \cup C$ = {SSF, SFS, FSS, SSS} = B,
$B \cap C$ = {SSF, SFS, SSS} = C

5. a. $\mathcal{S}$ = {(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 2, 1), (1, 2, 2), (1, 2, 3), (1, 3, 1), (1, 3, 2), (1, 3, 3), (2, 1, 1), (2, 1, 2), (2, 1, 3), (2, 2, 1), (2, 2, 2), (2, 2, 3), (2, 3, 1), (2, 3, 2), (2, 3, 3), (3, 1, 1), (3, 1, 2), (3, 1, 3), (3, 2, 1), (3, 2, 2), (3, 2, 3), (3, 3, 1), (3, 3, 2), (3, 3, 3)} **b.** {(1, 1, 1), (2, 2, 2), (3, 3, 3)} **c.** {(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)} **d.** {(1, 1, 1), (1, 1, 3), (1, 3, 1), (1, 3, 3), (3, 1, 1), (3, 1, 3), (3, 3, 1), (3, 3, 3)}

7. a. Há 35 resultados de $\mathcal{S}$. **b.** {*AABABAB, AABAABB, AAABBAB, AAABABB, AAAABBB*}

11. a. 0,07 **b.** 0,30 **c.** 0,57

13. a. 0,36 **b.** 0,64 **c.** 0,53
d. 0,47 **e.** 0,17 **f.** 0,75

15. a. 0,572 **b.** 0,879

17. a. Há pacotes de software estatísticos além do SPSS e do SAS.
b. 0,70 **c.** 0,80 **d.** 0,20

19. a. 0,8841 **b.** 0,0435

21. a. 0,10 **b.** 0,18; 0,19 **c.** 0,41 **d.** 0,59
e. 0,31 **f.** 0,69

23. a. 0,067 **b.** 0,400 **c.** 0,933 **d.** 0,533

25. a. 0,98 **b.** 0,02 **c.** 0,03 **d.** 0,24

27. a. 0,1 **b.** 0,7 **c.** 0,6

29. a. 20 **b.** 60 **c.** 10

31. a. 243 **b.** 3.645 dias (aproximadamente 10 anos)

33. a. 53,130 **b.** 1.190 **c.** 0,022 **d.** 0,023

35. 0,20

37. 0,0456

39. a. 0,0839 **b.** 0,24975

41. a. 0,929 **b.** 0,0714 **c.** 0,99997520

43. 0,000394; 0,00394; 0,00001539

45. a. 0,447; 0,500; 0,200 **b.** 0,400; 0,447 **c.** 0,211

47. a. 0,50 **b.** 0,50 **c.** 0,625
d. 0,375 **e.** 0,769

49. 0,217; 0,178

51. 0,436; 0,581

53. 0,083

55. a. 0,0111 **b.** 0,333 **c.** 0,2

59. a. 0,21 **b.** 0,455 **c.** 0,264; 0,274

61. a. 0,578; 0,278; 0,144 **b.** 0; 0,457; 0,543

63. b. 0,54 **c.** 0,68 **d.** 0,74 **e.** 0,7941

65. P(Média | S) = 0,3922, P(Mediana | S) = 0,2941, portanto Média e Mediana são mais e menos prováveis, respectivamente.

67. a. 0,126 **b.** 0,05 **c.** 0,1125
d. 0,2725 **e.** 0,5325 **f.** 0,2113

69. **a.** 0,300 **b.** 0,820 **c.** 0,146

73. 0,401; 0,722

75. **a.** 0,06235 **b.** 0,00421

77. 0,0059

79. **a.** 0,95

81. **a.** 0,10; 0,20 **b.** 0

83. **a.** $p(2 - p)$ **b.** $1 - (1 - p)^n$ **c.** $(1 - p)^3$
d. $0{,}9 + (1 - p)^3(0{,}1)$
e. $0{,}1(1 - p)^3/[0{,}9 + 0{,}1(1 - p)^3] = 0{,}0137$ para $p = 0{,}5$

85. 0,8588; 0,9897

87. $[2\pi(1 - \pi)]/(1 - \pi^2)$

89. **a.** 0,333; 0,444 **b.** 0,150 **c.** 0,291

91. 0,45; 0,32

93. **a.** 0,0083 **b.** 0,2 **c.** 0,2

95. 0,905

97. **a.** 0,956 **b.** 0,994

99. 0,926

101. **a.** 0,018 **b.** 0,601

103. **a.** 0,883; 0,117 **b.** 23 **c.** 0,156

105. $1 - (1 - p_1)(1 - p_2) \cdot \cdots \cdot (1 - p_n)$

107. **a.** 0,0417 **b.** 0,375

109. P(contratação do nº 1) = 6/24 para $s = 0$, = 11/24 para $s = 1$, = 10/24 para $s = 2$ e = 6/24 para $s = 3$, portanto $s = 1$ é melhor.

111. $1/4 = P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \cdot P(A_2) \cdot P(A_3) = 1/8$

Capítulo 3

1. $x = 0$ para FFF; $x = 1$ para SFF, FSF e FFS; $x = 2$ para SSF, SFS e FSS; e $x = 3$ para SSS

3. Z = média de dois números, com valores possíveis 2/2, 3/2, ... 12/2: W = valor absoluto da diferença, com valores possíveis 0, 1, 2, 3, 4, 5

5. Não. No Exemplo 3.4 assumimos $Y = 1$ se no máximo três baterias são examinadas e assumimos $Y = 0$ caso contrário. Então Y tem apenas dois valores.

7. **a.** {0, 1, ..., 12}; discreto **c.** {1, 2, 3, ...}; discreto **e.** {0, c, $2c$, ..., $10.000c$}, onde c é o direito autoral por livro; discreto **g.** $\{x: m \le x \le M\}$ onde m (M) é a tensão mínima (máxima) possível; contínua

9. **a.** {2, 4, 6, 8, ...}, ou seja, {2(1), 2(2), 2(3), 2(4), ...}, uma seqüência infinita; discreta
b. {2, 3, 4, 5, 6, ...}, ou seja, {1 + 1, 1 + 2, 1 + 3, 1 + 4, ...}, uma seqüência infinita; discreta

11. **a.** $p(4) = 0{,}45$, $p(6) = 0{,}40$, $p(8) = 0{,}15$, $p(x) = 0$ para $x \neq 4$, 6 ou 8 **c.** 0,55; 0,15

13. **a.** 0,70 **b.** 0,45 **c.** 0,55
d. 0,71 **e.** 0,65 **f.** 0,45

15. **a.** (1, 2), (1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5) **b.** $p(0) = 0{,}3$, $p(1) = 0{,}6$, $p(2) = 0{,}1$ **c.** $F(x) = 0$ para $x < 0$, = 0,3 para $0 \le x < 1$, = 0,9 para $1 \le x < 2$ e = 1 para $2 \le x$

17. **a.** 0,81 **b.** 0,162 **c.** É A; $AUUUA$, $UAUUA$, $UUAUA$, $UUUAA$; 0,00405

19. $p(1) = p(2) = p(3) = p(4) = 0{,}25$

21. $F(x) = 0$ para $x < 0$, = 0,10 para $0 \le x < 1$, = 0,25 para $1 \le x < 2$, = 0,45 para $2 \le x < 3$, = 0,70 para $3 \le x < 4$, = 0,90 para $4 \le x < 5$, = 0,96 para $5 \le x < 6$ e = 1 para $6 \le x$

23. **a.** $p(1) = 0{,}30$, $p(3) = 0{,}10$, $p(4) = 0{,}05$, $p(6) = 0{,}15$, $p(12) = 0{,}40$ **b.** 0,30, 0,60

25. **a.** $p(x) = (\frac{2}{3})^{x-1} \cdot (\frac{1}{3})$ para $x = 1, 2, 3, ...$
b. $p(y) = (\frac{2}{3})^{y-2} \cdot (\frac{1}{3})$ para $y = 2, 3, 4, 5, ...$
c. $p(z) = (\frac{25}{54}) \cdot (\frac{2}{3})^{2z-2}$ para $z = 1, 2, 3, ...$

29. **a.** 0,60 **b.** \$110

31. **a.** 16,38; 272,298; 3,9936 **b.** 401 **c.** 2.496
d. 13,66

33. Sim, porque $\sum_{x=1}^{\infty} 1/x^2 < \infty$.

35. \$700

37. $E(h(X)) = 0{,}408$, portanto joga.

39. $V(-X) = V(X)$

41. **a.** 32,5 **b.** 7,5
c. $V(X) = E[X(X - 1)] + E(X) - [E(X)]^2$

43. **a.** 0,25; 0,11; 0,06; 0,04; 0,01 **b.** 2,64; 1,54; 0,04, conservativo **c.** Eles são iguais **d.** $p(-1) = 0{,}02 = p(1)$, $p(0) = 0{,}96$

45. **a.** 0,850 **b.** 0,200 **c.** 0,200 **d.** 0,701
e. 0,851 **f.** 0,000 **g.** 0,570

47. **a.** 0,354 **b.** 0,115 **c.** 0,918

49. **a.** 0,403 **b.** 0,787 **c.** 0,774

51. 0,1478

53. 0,407, independência

55. **a.** 0,017 **b.** 0,811; 0,425 **c.** 0,006; 0,902; 0,586

57. Quando $p = 0{,}9$, a probabilidade é 0,99 para A e 0,9963 para B. Se $p = 0{,}5$, estas probabilidades são 0,75 e 0,6875, respectivamente

59. A tabulação para $p > 0{,}5$ é desnecessária

61. **a.** 20, 16 **b.** 70, 21

63. $P(|X - \mu| \ge 2\sigma) = 0{,}042$ quando $p = 0{,}5$ e = 0,065 quando $p = 0{,}75$, comparado ao limite superior de 0,25. Usando $k = 3$ no lugar de $k = 2$, estas probabilidades

são 0,002 e 0,004, respectivamente, considerando que o limite superior é 0,11.

65. **a.** 0,114 **b.** 0,879 **c.** 0,121 **d.** Use a distribuição com $n = 15, p = 0{,}10$

67. **a.** $h(x;\ 15, 10, 20)$ para $x = 5, ..., 10$
b. 0,0325 **c.** 0,697

69. **a.** $h(x;\ 10, 10, 20)$ **b.** 0,033

71. **a.** $nb(x;\ 2, 0{,}5)$ **b.** 0,188 **c.** 0,688 **d.** 2, 4

73. $nb(x;\ 6;\ 0{,}5)$, 6

75. **a.** 0,932 **b.** 0,065 **c.** 0,068
d. 0,492 **e.** 0,251

77. **a.** 0,011 **b.** 0,441 **c.** 0,554; 0,459 **d.** 0,945

79. Poisson(5) **a.** 0,492 **b.** 0,133

81. **a.** 0,313; 0,809; 0,283 **b.** 12; 3,464
c. 0,530; 0,011

83. **a.** 0,099 **b.** 0,135 **c.** 2

85. **a.** 4 **b.** 0,215 **c.** Pelo menos $-\ln(0{,}1)/2 \approx$ 1,1513 anos

87. **a.** 0,221 **b.** 6.800,000 **c.** $p(x;\ 20{,}106)$

91. **b.** 3,114; 0,405; 0,636

93. **a.** $b(x;\ 15;\ 0{,}75)$ **b.** 0,686
c. 0,313 **d.** 11,25; 2,81 **e.** 0,310

95. 0,991

97. **a.** $p(x;\ 2{,}5)$ **b.** 0,067 **c.** 0,109

99. 1,813; 3,05

101. $p(2) = p^2$, $p(3) = (1 - p)p^2$, $p(4) = (1 - p)p^2$, $p(x) = [1 - p(2) - >>> - p(x - 3)](1 - p)p^2$ para $x = 5, 6, 7, ...$; 0,99950841

103. **a.** 0,0029 **b.** 0,0767; 0,9702

105. **a.** 0,135 **b.** 0,00144 **c.** $\sum_{x=0}^{\infty}[p(x;\ 2)]^5$

107. 3,590

109. **a.** Não **b.** 0,0273

111. **b.** $0{,}6p(x;\ \lambda) + 0{,}4p(x;\ \mu)$ **c.** $(\lambda + \mu)/2$
d. $(\lambda - \mu)^2/4 + (\lambda + \mu)/2$

113. $\sum_{i=1}^{10}(p_{i+j+1} + p_{i-j-1})p_i$, onde $p_k = 0$ se $k < 0$ ou $k > 10$.

Capítulo 4

1. **a.** 0,25 **b.** 0,50 **c.** 0,4375

3. **b.** 0,5 **c.** 0,6875 **d.** 0,6328

5. **a.** 0,375 **b.** 0,125 **c.** 0,297 **d.** 0,578

7. **a.** $f(x) = 0{,}1$ para $25 \leq x \leq 35$ e 0 caso contrário
b. 0,20 **c.** 0,40 **d.** 0,20

9. **a.** 0,562 **b.** 0,438; 0,438 **c.** 0,071

11. **a.** 0,25 **b.** 0,1875 **c.** 0,9375 **d.** 1,4142
e. $f(x) = x/2$ para $0 < x < 2$
f. 1,33 **g.** 0,222; 0,471 **h.** 2

13. **a.** 3 **b.** 0 para $x \leq 1$, $1 - x^{-3}$ para $x > 1$
c. 0,125; 0,088 **d.** 1,5; 0,866 **e.** 0,924

15. **a.** $F(x) = 0$ para $x \leq 0$, $= 90[\frac{x^9}{9} - \frac{x^{10}}{10}]$ para $0 < x < 1$, $=$ 1 para $x \geq 1$ **b.** 0,0107 **c.** 0,0107; 0,0107
d. 0,9036 **e.** 0,818; 0,111 **f.** 0,3137

17. **a.** Para $2 \leq x \leq 4$, $F(x) = 0{,}25[3x - 7 - (x - 3)^3]$
b. 3 **c.** 3; 0,2

19. **a.** 0,597 **b.** 0,369
c. $f(x) = 0{,}3466 - 0{,}25\ln(x)$ para $0 < x < 4$

21. 314,79

23. 248; 3,60

25. **b.** 1,8(o nonagésimo percentil para X) + 32
c. a(X percentil) + b

27. **a.** 2,14 **b.** 0,81 **c.** 1,17
d. 0,97 **e.** 2,41

29. **a.** 2,54 **b.** 1,34 **c.** −0,42

31. **a.** 0,9918 **b.** 0,0082 **c.** 0,9544

33. **a.** 0,3336 **b.** Aproximadamente 0
c. 0,5795 **d.** 6,524 **e.** 0,8028

35. **a.** 36,7 **b.** 22,225 **c.** 3,179

37. 0,002

39. 10; 0,2

41. 7,3%

43. 21,155

45. **a.** 0,1190; 0,6969 **b.** 0,0021 **c.** 0,7054
d. >5020 ou <1844 (usando $z_{0,0005} = 3{,}295$)
e. Normal, $\mu = 7{,}576$, $\sigma = 1{,}064$; 0,7054

47. 0,3174 para $k = 1$; 0,0456 para $k = 2$; 0,0026 para $k = 3$, quando comparado com os limites de 1,25 e 0,111, respectivamente.

49. **a.** Exato: 0,212, 0,577, 0,573; Aproximado: 0,211, 0,567; 0,596 **b.** Exato: 0,885; 0,575; 0,017; Aproximado: 0,885; 0,579; 0,012 **c.** Exato: 0,002; 0,029; 0,617; Aproximado: 0,003; 0,033; 0,599

51. **a.** 0,9666 **b.** 0,0099; 0

53. **b.** Normal, $\mu = 239$, $\sigma^2 = 12{,}96$

55. **a.** 120 **b.** 1,329 **c.** 0,371
d. 0,735 **e.** 0

57. **a.** 5, 4 **b.** 0,715 **c.** 0,411

59. **a.** 1 **b.** 1 **c.** 0,982 **d.** 0,129

61. **a.** 0,449; 0,699; 0,148 **b.** 0,05; 0,018

63. **a.** $\cap A_i$ **b.** Exponencial com $\lambda = 0{,}05$
c. Exponencial com parâmetro $n\lambda$

67. **a.** 0,826; 0,826; 0,0636 **b.** 0,664 **c.** 172,727

71. **a.** 0,930 **b.** 0,298 **c.** 98,18

73. **a.** 68,0; 122,1 **b.** 0,3204
c. 0,7257, inclinação

75. **a.** 149,157, 223,595 **b.** 0,9573 **c.** 0,0414
d. 148,41 **e.** 9,57 **f.** 125,90

77. $\alpha = \beta$

79. **b.** $[\Gamma(\alpha + \beta) \cdot \Gamma(m + \beta)]/[\Gamma(\alpha + \beta + m) \cdot \Gamma(\beta)]$, $\beta/(\alpha + \beta)$

81. Sim, visto que o padrão na marcação é bem linear.

83. Sim

85. Sim

87. Marcação ln(x) *versus* percentil z. O padrão é retilíneo, portanto uma distribuição de população lognormal é plausível.

89. O padrão na marcação é bem linear; é muito plausível que a força seja distribuída normalmente.

91. Há curvatura considerável na marcação. λ é um parâmetro de escala (como é σ para a família normal).

93. **a.** $F(y) = \frac{1}{48}(y^2 - y^3/18)$ para $0 \leq y \leq 12$
b. 0,259; 0,5; 0,241 **c.** 6; 43,2; 7,2
d. 0,518 **e.** 3,75

95. **a.** $f(x) = x^2$ para $0 \leq x < 1$ e $= \frac{7}{4} - \frac{3}{4}x$ para $1 \leq x \leq \frac{7}{3}$
b. 0,917 **c.** 1,213

97. **a.** 0,9162 **b.** 0,9549 **c.** 1,3374

99. **a.** 0,3859 **b.** 0,0663 **c.** (72,97; 119,03)

101. **b.** $F(x) = 0$ para $x < -1$, $= (4x - x^3/3)/9 + \frac{11}{27}$ para $-1 \leq x \leq 2$ e $= 1$ para $x > 2$
c. Não. $F(0) < 0{,}5 \Rightarrow \tilde{\mu} > 0$
d. $Y \sim \text{Bin}(10, \frac{5}{27})$

103. **a.** 0,368; 0,828; 0,460
b. 352,53
c. $1/\beta \cdot \exp[-\exp(-(x - \alpha)/\beta)] \cdot \exp(-(x - \alpha)/\beta)$
d. α
e. $\mu = 201{,}95$, moda $= 150$, $\tilde{\mu} = 182{,}99$

105. **a.** μ **b.** Não **c.** 0
d. $(\alpha - 1)\beta$ **e.** $\nu - 2$

107. **b.** $p(1 - \exp(-\lambda_1 x)) + (1 - p)(1 - \exp(-\lambda_2 x))$ para $x \geq 0$ **c.** $p/\lambda_1 + (1 - p)/\lambda_2$
d. $V(X) = 2p/\lambda_1^2 + 2(1 - p)/\lambda_2^2 - \mu^2$
e. 1, $CV > 1$ **f.** $CV < 1$

109. **a.** Lognormal **b.** 1 **c.** 2,72; 0,0185

113. **a.** Exponencial com $\lambda = 1$
c. Gama com parâmetros α e $c\beta$

115. **a.** $(1/365)^3$ **b.** $(1/365)^2$ **c.** 0,000002145

117. **b.** Sejam $u_1, u_2, u_3, \ldots$ uma seqüência de observações de uma distribuição Unif[0,1] (uma seqüência de números aleatórios). Então com $x_i = (-0{,}1)\ln(1 - u_i)$, os x_is são observações de uma distribuição exponencial com $\lambda = 10$.

119. $g(E(X)) \leq E(g(X))$

Capítulo 5

1. **a.** 0,20 **b.** 0,42 **c.** Pelo menos uma mangueira é utilizada em cada bomba: 0,70. **d.** $p_X(x) = 0{,}16$; 0,34; 0,50 para $x = 0, 1, 2$, respectivamente; $p_Y(y) = 0{,}24$; 0,38; 0,38 para $y = 0, 1, 2$, respectivamente; 0,50
e. Não; $p(0, 0) \neq p_X(0) \cdot p_Y(0)$

3. **a.** 0,15 **b.** 0,40 **c.** 0,22 **d.** 0,17; 0,46

5. **a.** 0,054 **b.** 0,00018

7. **a.** 0,030 **b.** 0,120 **c.** 0,300
d. 0,380 **e.** Sim

9. **a.** 3/380,000 **b.** 0,3024 **c.** 0,3593
d. $10Kx^2 + 0{,}05$ para $20 \leq x \leq 30$ **e.** Não

11. **a.** $e^{-\lambda-\mu} \cdot \lambda^x \cdot \mu^y/x!y!$ **b.** $e^{-\lambda-\mu} \cdot [1 + \lambda + \mu]$
c. $e^{-(\lambda+\mu)} \cdot (\lambda + \mu)^m/m!$; Poisson $(\lambda + \mu)$

13. **a.** e^{-x-y} para $x \geq 0$, $y \geq 0$ **b.** 0,400 **c.** 0,594
d. 0,330

15. **a.** $F(y) = 1 - e^{-\lambda y} + (1 - e^{-\lambda y})^2 - (1 - e^{-\lambda y})^3$ para $y \geq 0$
b. $2/3\lambda$

17. **a.** 0,25 **b.** 0,318 **c.** 0,637
d. $f_X(x) = 2\sqrt{R^2 - x^2}/\pi R^2$ para $-R \leq x \leq R$; não

19. **a.** $K(x^2 + y^2)/(10Kx^2 + 0{,}05)$; $K(x^2 + y^2)/(10Ky^2 + 0{,}05)$
b. 0,556; 0,549 **c.** 25,37; 2,87

21. $= f_Y(y)$ para todo x

23. 0,15

25. L^2

27. $\frac{1}{3}$ h

29. $-\frac{2}{3}$

31. **a.** −0,1082 **b.** −0,0131

37. **a.**

$\bar{x}$	25	32,5	40	45	52,5	65
$p(\bar{x})$	0,04	0,20	0,25	0,12	0,30	0,09

, $E(\bar{X}) = \mu = 44{,}5$

b.

s^2	0	112,5	312,5	800
$p(s^2)$	0,38	0,20	0,30	0,12

, $E(S^2) = 212{,}25 = \sigma^2$

39.

Proporção	0	0,1	0,2	0,3	0,4	0,5
Probabilidade	0,000	0,000	0,000	0,001	0,005	0,027

Proporção	0,6	0,7	0,8	0,9	1,0
Probabilidade	0,088	0,201	0,302	0,269	0,107

41. **a.**

$\bar{x}$	1	1,5	2	2,5	3	3,5	4
$p(\bar{x})$	0,16	0,24	0,25	0,20	0,10	0,04	0,01

b. 0,85 **c.**

r	0	1	2	3
$p(r)$	0,30	0,40	0,22	0,08

47. a. 0,6826 **b.** 0,1056

49. a. 0,6026 **b.** 0,2981

51. 0,7720

53. a. 0,0062 **b.** 0

55. a. 0,0968 **b.** 0,8882

57. 0,9616

59. a. 0,9986; 0,9986 **b.** 0,9015; 0,3970
c. 0,8357 **d.** 0,9525; 0,0003

61. a. 3,5; 2,27; 1,51 **b.** 15,4; 75,94; 8,71

63. a. 0,695 **b.** 4,0675 > 2,6775

65. a. 0,9232 **b.** 0,9660

67. 0,1588

69. a. 2.400 **b.** 1.205; independência
c. 2.400; 41,77

71. a. 158; 430,25 **b.** 0,9788

73. a. Aproximadamente normal com média = 105, DP = 1,2649; Aproximadamente normal com média = 100, DP = 1,0142
b. Aproximadamente normal com média = 5, DP = 1,6213
c. 0,0068 **d.** 0,0010, sim

75. a. 0,2; 0,5; 0,3 para x = 12, 15, 20; 0,10; 0,35; 0,55 para y = 12, 15, 20 **b.** 0,25 **c.** Não **d.** 33,35 **e.** 3,85

77. a. 3/81,250 **b.** $f_X(x) = k(250x - 10x^2)$ para $0 \leq x \leq 20$ e $= k(450x - 30x^2 + \frac{1}{2}x^3)$ para $20 < x \leq 30$; $f_Y(y)$ resulta da substituição de y por x em $f_X(x)$. Eles não são independentes.
c. 0,355 **d.** 25,969
e. 204,6154; −0,894 **f.** 7,66

79. ≈ 1

81. a. 400 min **b.** 70

83. 97

85. 0,9973

89. *b, c.* Qui-quadrado com $\nu = n$.

91. a. $\sigma_W^2/(\sigma_W^2 + \sigma_E^2)$ **b.** 0,9999

93. 26; 1,64

Capítulo 6

1. a. 8,14, $\overline{X}$ **b.** 0,77, $\tilde{X}$ **c.** 1,66, S
d. 0,148 **e.** 0,204, $S/\overline{X}$

3. a. 1,348, $\overline{X}$ **b.** 1,348, $\overline{X}$
c. 1,781, $\overline{X} + 1{,}28S$
d. 0,6736 **e.** 0,0905

5. 1,703,000; 1,591,300; 1,601,438,281

7. a. 120,6 **b.** 1.206.000 **c.** 0,80 **d.** 120,0

9. a. 2,11 **b.** 0,119

11. b. $\left[\dfrac{p_1q_1}{n_1} + \dfrac{p_2q_2}{n_2}\right]^{1/2}$
c. Use $\hat{p}_i = x_i/n_i$ e $\hat{q}_i = 1 - \hat{p}_i$ no lugar de p_i e q_i na parte (b) para i = 1, 2.
d. −0,245 **e.** 0,041

15. a. $\hat{\theta} = \sum X_i^2/2n$ **b.** 74,505

17. b. 0,444

19. a. $\hat{p} = 2\hat{\lambda} - 0{,}30 = 0{,}20$ **b.** $\hat{p} = (100\hat{\lambda} - 9)/70$

21. b. $\hat{\alpha} = 5$, $\hat{\beta} = 28{,}0/\Gamma(1{,}2)$

23. $\hat{\lambda}_1 = \bar{x}$, $\hat{\lambda}_2 = \bar{y}$, estimado de $(\lambda_1 - \lambda_2)$ é $\bar{x} - \bar{y}$.

25. a. 384,4; 18,86 **b.** 415,42

29. a. $\hat{\theta} = \text{mín}(X_i)$, $\hat{\lambda} = n/\sum[X_i - \text{mín}(X_i)]$
b. 0,64; 0,202

33. Com x_i = tempo entre o nascimento $i - 1$ e o nascimento i, $\hat{\lambda} = 6/\sum_{i=1}^{6} ix_i = 0{,}0436$.

35. 29,5

37. 1,0132

Capítulo 7

1. a. 99,5% **b.** 85% **c.** 2,96 **d.** 1,15

3. a. Mais estreito **b.** Não **c.** Não **d.** Não

5. a. (4,52; 5,18) **b.** (4,12; 5,00)
c. 0,55 **d.** 94

7. Por um fator de 4; a largura é decrescente por um fator de 5.

9. a. $(\bar{x} - 1{,}645\sigma/\sqrt{n}, \infty)$; $(4{,}57, \infty)$
b. $(\bar{x} - z_\alpha \cdot \sigma/\sqrt{n}, \infty)$ **c.** $(-\infty, \bar{x} + z_\alpha \cdot \sigma/\sqrt{n})$; $(-\infty, 59{,}7)$

11. 950; 0,8714

13. a. (0,990; 1,066) **b.** 158

15. a. 80% **b.** 98% **c.** 75%

17. 134,53

19. (0,513; 0,615)

21. 0,218

23. a. (0,438; 0,814) **b.** 659

25. a. 381 **b.** 339

29. a. 2,228 **b.** 2,086 **c.** 2,845 **d.** 2,680
e. 2,485 **f.** 2,571

31. a. 1,812 **b.** 1,753 **c.** 2,602 **d.** 3,747
e. 2,1716 (do MINITAB) **f.** Aproximadamente 2,43

33. a. Quantidade razoável de simetria, sem outliers
b. Sim (baseado em uma marcação de probabilidade normal)
c. (430,5; 446,1), sim, não

35. a. 95% CI: (2783,3; 2991,9)
b. 95% PI: (2632,1; 3143,1), aproximadamente 2,5 vezes mais largo

37. a. (0,888; 0,964) **b.** (0,752; 1,100)
c. (0,634; 1,218)

39. a. Sim **b.** (6,45; 98,01)
c. (18,63; 85,83)

41. Todos os 70%; (c), porque é o mais curto

43. a. 18,307 **b.** 3,940 **c.** 0,95 **d.** 0,10

45. (3,6; 8,1); não

47. a. 95% IC: (6,702; 9,456) **b.** (0,166; 0,410)

49. a. Parece haver uma ligeira inclinação positiva na metade do meio da amostra, mas as suíças inferiores são mais longas do que as superiores. A extensão da variabilidade é muito substancial, embora não haja outliers.
b. Sim. O padrão dos pontos em uma marcação de probabilidade normal é razoavelmente linear.
c. (33,53; 43,79)

51. a. (0,624; 0,732) **b.** 1.080 **c.** Não

53. (−0,84, −0,16)

55. 246

57. $(2t_r/\chi^2_{1-\alpha/2,2r}, 2t_r/\chi^2_{\alpha/2,2r}) = (65{,}3; 232{,}5)$

59. a. $(\text{máx}(x_i)/(1-\alpha/2)^{1/n}, \text{máx}(x_i)/(\alpha/2)^{1/n})$
b. $(\text{máx}(x_i), \text{máx}(x_i)/\alpha^{1/n})$ **c.** (b); (4,2; 7,65)

61. (73,6; 78,8) versus (75,1; 79,6)

Capítulo 8

1. a. Sim **b.** Não **c.** Não
d. Sim **e.** Não **f.** Sim

5. H_0: $\sigma = 0{,}05$ versus H_a: $\sigma < 0{,}05$. I: conclui que a variabilidade na espessura é satisfatória quando de fato não é. II: conclui que a variabilidade na espessura não é satisfatória quando de fato é.

7. I: concluindo que a usina não está em conformidade quando está; II: concluindo que a usina está em conformidade quando não está.

9. a. R_1 **b.** I: julgando que uma das duas companhias é favorecida em relação à outra quando este não é o caso: II: julgando que nenhuma companhia é favorecida em relação a outra quando na realidade uma das duas é realmente a preferida. **c.** 0,044
d. $\beta(0{,}3) = \beta(0{,}7) = 0{,}488$, $\beta(0{,}4) = \beta(0{,}6) = 0{,}845$
e. Rejeitar H_0 em favor de H_a.

11. a. H_0: $\mu = 10$ versus H_a: $\mu \neq 10$ **b.** 0,01
c. 0,5319; 0,0078 **d.** 2,58
e. 10,1032 é substituído por 10,124, e 9,8968 é substituído por 9,876. **f.** $\bar{x} = 10{,}020$, portanto H_0 não deve ser rejeitada. **g.** $z \geq 2{,}58$ ou $\leq -2{,}58$

13. b. 0,0004; 0; menor do que 0,01

15. a. 0,0301 **b.** 0,003 **c.** 0,004

17. a. $z = 2{,}56 \geq 2{,}33$; portanto rejeita H_0.
b. 0,8413 **c.** 143 **d.** 0,0052

19. a. $z = -2{,}27$; portanto não rejeita H_0. **b.** 0,2266
c. 22

21. a. $t_{0{,}025,12} = 2{,}179 > 1{,}6$; portanto não rejeita H_0: $\mu = 0{,}5$.
b. $-1{,}6 > -2{,}179$; portanto não rejeita H_0.
c. Não rejeita H_0.
d. Rejeita H_0 em favor de H_a: $\mu \neq 0{,}5$.

23. $t = 2{,}24 \geq 1{,}708$; portanto H_0 deve ser rejeitada. Os dados sugerem uma contradição da primeira hipótese.

25. a. $z = -3{,}33 \leq -2{,}58$; portanto rejeita H_0.
b. 0,1056 **c.** 217

27. $-2{,}09 > -2{,}33$; portanto não rejeita H_0.

29. a. $0{,}498 < 1{,}895$; portanto não rejeita H_0. **b.** 0,72

31. $-1{,}24 > -1{,}397$; portanto a primeira hipótese não parece ter sido contradita

35. Sim, porque $-2{,}47 \leq -1{,}96$.

37. $z = 3{,}67 \geq 2{,}58$; portanto rejeita H_0: $p = 0{,}40$. Não.

39. −6,1 é muito menor do que −2,33; portanto os dados suportam fortemente esta conclusão.

41. a. {15, ..., 20} **b.** 0,021, sim **c.** 0,874; 0,196
d. Usando $R = \{14, ..., 20\}$, não.

43. $z = -0{,}61 > -2{,}33$; não

45. a. Não rejeita H_0. **b.** Não rejeita H_0.
c. Não rejeita H_0. **d.** Rejeita H_0.
e. Não rejeita H_0. **f.** Não rejeita H_0.

47. a. 0,0358 **b.** 0,0802 **c.** 0,5824
d. 0,1586 **e.** 0

49. a. Valor $P = 0{,}003 < 0{,}05$; portanto rejeita H_0.
b. Valor $P = 0{,}055 > 0{,}01$; portanto H_0 não pode ser rejeitada.
c. Valor $P > 0{,}5$; portanto H_0 não pode ser rejeitada em qualquer α razoável.